ПСИХОЛОГИЧЕСКИЙ ЖУРНАЛ, 2014, том 35, № 6, с. 99–113

МЕТОДЫ И МЕТОДИКИ

НАДЕЖНОСТЬ И ДОСТОВЕРНОСТЬ В КОНТЕНТ-АНАЛИЗЕ ТЕКСТОВ: ВЫБОР ПОКАЗАТЕЛЕЙ

© 2014 г. А. Н. Олейник*, И. П. Попова, С. Г. Кирдина***, Т. Ю. Шаталова******

** Доктор экономических наук, PhD, ведущий научный сотрудник Центрального экономико-математического института РАН, Университет Мемориал (Канада); e-mail: aoleynik@ mun.ca*

*** Кандидат социологических наук, старший научный сотрудник Института социологии РАН, ведущий научный сотрудник НИУ – Высшая школа экономики; e-mail: irina_popova@list.ru*

**** Доктор социологических наук, зав. сектором Института экономики РАН; e-mail: kirdina777@gmail.com*

***** преподаватель Московского государственного университета имени М.В. Ломоносова; e-mail: shatalova-tanya@yandex.ru.*

Рассматривается контент-анализ как один из методов анализа первичных текстовых данных, применяемый в психологических, социологических и других исследованиях. Обсуждаются такие показатели его надежности, как π (*пи*) Скотта, α (*альфа*) Криппендорфа, *S* Беннета, *к* (*каппа*) Коэна, *I* Перро и Лея, а также допущения, используемые при их расчетах. Показано, что данный список может быть дополнен, с одной стороны, коэффициентами корреляции между распределением качественных кодов и совместной сочетаемости слов, а с другой – между распределением качественных кодов и распределением категорий, полученных с помощью словаря, основанного на замещении. Показано, что выбор конкретного показателя надежности зависит от текста, а именно: его формата (стилистический или риторический) и типа прочтения (понимание или интерпретация). В частности, *к* и *S* представляются уместными для зависящего от конкретной точки зрения анализа, например, при прочтении риторических текстов. Представлены результаты контент-анализа 57 текстов, осуществленного четырьмя кодировщиками с помощью специализированной компьютерной программы *QDA Miner*.

Ключевые слова: контент-анализ, показатели надежности, корреляционный анализ, интерпретация, понимание, стилистические тексты, риторические тексты, психология чтения

Оценка любого исследования с точки зрения критериев достоверности и надежности представляет собой важный шаг в его проектировании, внедрении и распространении результатов. Соответствие между инструментарием исследования или исследовательскими данными, с одной стороны, и изучаемым явлением, с другой стороны, увеличивает его достоверность. В свою очередь, надежность зависит от стабильности инструментария исследования или его результатов во времени и/или при одновременном использовании инструментария несколькими исследователями. Существует несколько видов надежности в зависимости от занимаемой эпистемологической позиции (позитивистской или объяснительной/интерпретативной) и от характеристики программы исследований: основными критериями могут выступать такие виды надежности, как надежность-стабильность, надежность-воспроизводимость и надежность-точность.

Контент-анализ является одним из методов анализа и обработки первичных данных, применяемым в целом ряде социальных наук, включая психологию. Он предназначен для описания, анализа и интерпретации смыслов, которые содержатся в текстах или изображениях. Диапазон его использования – от кодирования открытых вопросов в массовых опросах населения [42, 32] и оценки согласия между экспертами в психологических и социологических исследованиях – до создания аннотаций в лингвистике и библиографических исследованиях [23], включая подготовку обзоров научной литературы [7].

В психологии можно выделить четыре основные области применения контент-анализа [30, с. 11–12]. Во-первых, это исследование транс-

криптов интервью для выявления мотивационных, ментальных и личностных характеристик человека. Классические работы Олпорта [21], Болдуина [24] и Уайта [47] служат наиболее известными примерами таких исследований. Во-вторых, при обработке ответов на открытые вопросы интервью и различных тестов, таких как Тематический апперцептивный тест (*Thematic Apperception Test*) – ТАТ. В-третьих, это исследования коммуникационных процессов при условии признания значимым их содержания [25], например, психологические теории чтения, к обсуждению которых мы еще вернемся в настоящей статье. В-четвертых, контент-анализ полезен при обобщении терминов, с помощью которых испытуемые осмысляют различные ситуации или культурные контексты. Так, шкала (семантический дифференциал) Осгуда была адаптирована для сравнения культур [38, 39]. Среди российских психологов, применяющих контент-анализ в своих исследованиях, можно назвать Т.Н. Ушакову, С.С. Белову и Е.А. Валуеву [20], Н.Д. Павлову [12], Е.А. Володарскую и Н.Ю. Логвинову [4], М.И. Воловикову и Л.М. Соснину [5], работы которых близки к четвертой обозначенной выше области. В свою очередь, работы Н.Е. Лысенко и Д.М. Давыдова [9], В.В. Абраменковой [1] и О.А. Гулевич [6] можно отнести ко второй области, когда речь идет об анализе ответов на интервью и разработке тестов.

При оценке исследований с использованием контент-анализа особое внимание требуется уделять такой группе критериев, как надежность-воспроизводимость. Так, при кодировании данных суждения кодировщика могут быть субъективными. Привлечение нескольких кодировщиков повышает объективность анализа, но требует измерения степени согласия между ними. Для таких случаев Криппендорф определяет надежность-воспроизводимость как “степень, в которой процесс может быть повторен различными исследователями, работающими в самых разных условиях, в разных местах или с использованием различных, но функционально эквивалентных инструментов” [30, с. 215].

Контент-аналитики имеют в своем распоряжении несколько показателей надежности-воспроизводимости: *S* Беннета и соавторов, π Скотта, κ Коэна, α Криппендорфа – лишь немногие из них. Такой плюрализм создает некоторую путаницу, ставя вопрос о замещении или дополнении этими показателями друг друга. Если имеет место замещение, то в таком случае исследователь определяет лучший показатель надежности, который “превосходит” другие. Утверждение о том, что “из существующих показателей α Криппендорфа лучше всего подходит в качестве стандарта” [29, с. 78], иллюстрирует такой подход. Если же имеет место взаимодополняемость, исследователь пытается разграничить области применения существующих показателей. В этом случае необходимо определить, при каких условиях (параметры текстов, которые подлежат анализу; количество кодировщиков; число категорий, включенных в книгу кодов и др.) более целесообразно использование α Криппендорфа, а не, например, *S* Беннета и соавторов.

Цель статьи – представить дополнительные аргументы и эмпирических данные в поддержку предположения о взаимодополняющем характере методов измерения надежности. В отличие от других работ, посвященных данной проблеме [23, с. 586–591; 32, с. 530–533], представленное ниже исследование было направлено на то, чтобы предварительно разграничить области применения наиболее популярных показателей надежности: во-первых, в соответствии с особенностями текста, подлежащего контент-анализу (*стилистический* или *риторический*), и, во-вторых, в зависимости от конкретной цели контент-анализа (*понимание* авторского сообщения в отличие от читательской *интерпретации*).

В первой части статьи дается краткий обзор самых популярных показателей надежности-воспроизводимости. Обсуждение допущений, на которых основаны те или иные показатели надежности, помогает уточнить вопрос исследования: как контекст контент-анализа влияет на их применимость. Во второй части рассматриваются данные проведенного исследования по изучению академического чтения – того, как ученые читают и понимают работы друг друга. В третьей части эмпирически проверяются две гипотезы относительно выбора показателей надежности.

АНАЛИЗ ИМЕЮЩИХСЯ ПОКАЗАТЕЛЕЙ

Ознакомиться с всесторонним обзором существующих показателей надежности-воспроизводимости, включая обсуждение математических оснований их расчета, можно в других работах [30, 34]. Наша задача в данном случае более ограничена. Она предполагает выявление различий в допущениях, на которых основаны показатели надежности.

Различия в допущениях часто остаются без должного внимания из-за того факта, что большинство известных индексов надежности-воспроизводимости следует общей логике. Оставляя

в стороне простой расчет процента согласия (т.е. процента единиц, одинаково закодированных всеми кодировщиками), расчет показателей надежности предусматривает сравнение наблюдаемого процента согласия или несогласия с ожидаемым. Другими словами, достигнутый уровень (не)согласия сравнивается с уровнем (не)согласия, который может быть получен случайно. "Значение $1-A_e$ [A означает согласие] показывает максимально возможную величину согласия за вычетом фактора случайности; значение A_o-A_e показывает наблюдаемую величину согласия за вычетом фактора случайности. Далее, соотношение между A_o-A_e и $1-A_e$ говорит нам, какая пропорция возможного неслучайного согласия в действительности наблюдалась. Эта идея выражается следующей формулой: S, π, $\kappa = \frac{A_o - A_e}{1 - A_e}$" [23, с. 559]. "$\alpha$ имеет такое же логическое обоснование, но для ее расчета нужно учитывать уровни несогласия: D_o наблюдаемого и D_e ожидаемого" [30, с. 248].

Показатели надежности начинают расходиться с момента расчета ожидаемого (не)согласия, A_e. На первый взгляд, это простая формальность, однако она позволяет увидеть различия в допущениях, на которых эти показатели основаны.

Показатель π Скотта. Данный критерий корректирует процент согласия "на количество категорий в коде и частоты, с которой каждая категория использована" [42, с. 323]. Это достигается путем сравнения наблюдаемого распределения категорий с ожидаемым, как в общем случае.

Однако Скотт делает очень специфическое предположение относительно соответствия между наблюдаемым и ожидаемым распределениями. Предполагается, что наблюдаемое распределение категорий кодировщиками между фрагментами текста указывает на их "истинное" распределение, которое служит в качестве основы для расчета ожидаемого распределения. Иными словами, π использует "реальное поведение кодировщиков для оценки предшествующего кодированию распределения категорий" [23, с. 561].

Предположение, сделанное Скоттом, означает, что произвольное (ожидаемое) распределение категорий между кодируемыми единицами (текстов) любым кодировщиком регулируется распределением категорий в реальном мире. Отсюда следует, что подготовленные и компетентные кодировщики, как правило, идентифицируют все единицы, которые соответствуют определенной категории. Если один кодировщик проглядит релевантную единицу, то второй, работающий независимо от первого, но использующий ту же книгу кодов, вероятно, заметит пропущенную, и наоборот[1]. Единицы, идентифицированные совместными усилиями кодировщиков, соответствуют их первоначальному, "истинному" распределению, которое существует независимо от кодировщиков. Также принимается допущение, что кодировщики взаимозаменяемы и обладают одинаковой квалификацией. Чем больше кодировщиков включено в процесс кодирования, тем меньше шансов для упущений. Исходная формула для расчета π в случае двух кодировщиков впоследствии была обобщена для случаев, где их больше двух: расчетом π для каждой пары кодировщиков и их суммированием [32, с. 526].

Показатель α Криппендорфа. Критерий, рассматриваемый в данном разделе, и показатель π Скотта получены в результате сходных допущений. В случае α параметры совокупности ("истинное" распределение категорий) приблизительно подсчитаны, исходя из эмпирически наблюдаемых соотношений (отражаемых в итоговых строке и столбце таблицы сопряженности). Криппендорф иллюстрирует это утверждение рассмотрением примера двух кодировщиков, Джона и Хана, которые должны идентифицировать особую категорию статей (статьи на китайском языке, в которых упоминаются США). "Имея найденные Джоном 2 ссылки на США из 10 статей и идентифицированные Ханом 4 статьи из 10, эти два наблюдателя совместными усилиями идентифицировали 6 из 20… это наша оценка параметров совокупности" [30, с. 225]. Параметры статей со ссылками на США не зависят ни от того, являются ли они объектом контент-анализа, ни от его результатов.

Этот простой пример ясно показывает, что "истинное" распределение категорий предшествует кодированию. Предположение, что параметры совокупности заранее известны, может быть обоснованным в некоторых областях исследований, например, в медицине и медицинской психологии, где квалифицированные кодировщики обычно имеют представления о соотношении единиц, которые подходят под описание определенных категорий (как в случае определения болезни на основе наблюдаемых симптомов). В этом случае результаты контент-анализа выборки текстов или изображений помогают рассчитать параметры. Оценка параметров совокупности затем вво-

[1] Аналогичное предположение лежит в основе использования α Кронбаха в теории культурного консенсуса. Согласие между кодировщиками предположительно зависит от того, как хорошо они знают контекст культурной сферы, который существует независимо от их усилий по кодированию [46, с. 343].

дится в расчеты ожидаемого уровня несогласия между кодировщиками.

Эмпирически наблюдаемое соотношение представляет собой ключевой компонент формулы для расчета α: $\alpha = 1 - \frac{n-1}{n}(b+c)/2\bar{p}\bar{q}$, где:

– n – число значений, использованных совместно обоими кодировщиками ($n = 2N$, количество единиц в таблице сопряженности),

– $(b+c)$ – пропорция случаев, когда кодировщики не согласны – их решения не совпадают (клетки вне диагонали в таблице сопряженности),

– $\bar{p}\bar{q}$ – оценка совокупности (итоговая строка и столбец таблицы сопряженности) [30, с. 248].

Незначимость количества кодировщиков и количества категорий для расчета ожидаемого распределения означают, что эти переменные не учитываются явным образом.

Как и в случае π, расчет α содержит предположение, что кодировщики взаимозаменяемы[2]. Согласно сторонникам использования α, хороший показатель надежности "должен быть (а) независимым от количества занятых наблюдателей и (б) инвариантным в отношении перестановок и избирательного участия наблюдателей. При этих двух условиях согласие не будет находиться под влиянием индивидуальных идентичностей и количества наблюдателей, которым выпало принять участие в кодировании" [29, с. 79]. Некоторые эмпирические данные действительно поддерживают утверждение о том, что количество кодировщиков не влияет на величину α [32, с. 530].

Тот же источник, однако, предполагает, что α зависит от количества категорий, включенных в книгу кодов, и это нарушает стабильность и точность данного показателя по мере роста числа категорий. Другая отмечаемая ограниченность α относится к ситуации преобладания отдельных категорий в данных. Когда превалируют одна-две категории, кодировщики могут быть согласными в большинстве случаев, что будет соответствовать реальности, но значение α все же останется низким [23, с. 573].

Показатель κ Коэна. В отличие от π и α, расчет κ предполагает использование иной точки отсчета. Степень случайного согласия интерпретируется в терминах последовательности кодировщика в категоризации единиц анализа [23, с. 561, 570]. С этой точки зрения, итоговая строка и столбец в таблице сопряженности указывают на индивидуальные предпочтения и предубеждения кодировщиков [40, с. 139], а не на действительное распределение единиц среди категорий, как в случае π и α. Первоначально κ была рассчитана для случая двух кодировщиков. Ее формула позже была обобщена для случая множества кодировщиков методом агрегирования попарных коэффициентов согласия [43, с. 285; 23, с. 562; 28, с. 763].

В данном случае основное предположение заключается в том, что "вероятность того, что объект отнесен к определенной категории, в целом у кодировщиков не варьирует" [43, с. 291]. Другими словами, кодировщики согласны не потому, что их решения определены "невидимой рукой" "истинного" распределения категорий, а потому, что имеют сходные мнения или сходные пристрастия. Их решения имеют свойство быть субъективными и психологически обусловленными, что не мешает им достигать согласия. Это предположение оказывается оправданным прежде всего в парадигмальных науках. Поэтому неудивительно замечание о том, что разрабатывая свой критерий и думая о его использовании, Коэн "заботился главным образом о применении в психологии, где зачастую есть предшествующее знание о вероятном распределении наблюдений в ячейках" [40, с. 139].

Изменение точки отсчета с "истинного", естественным образом возникающего распределения категорий, на субъективные суждения кодировщиков привело к жесткому отпору со стороны апологетов π и α. Они полагают, что "κ относится к двум индивидуальным наблюдателям, а не к совокупности данных, которые они рассматривают, которая и должна быть предметом дискуссий о надежности" ([30, с. 248]; см. также [29, с. 81]). Тем не менее, κ и α могут относиться к различным аспектам надежности: если первая в большей мере характеризует надежность суждений, то α Криппендорфа касается в основном надежности данных.

Показатель *S* Беннета. Этот критерий Беннета и соавторов представляет собой другой пример коэффициента надежности, когда за точку отсчета используется субъективное суждение кодировщи-

[2] Это предположение позволяет также минимизировать влияние ценностей кодировщиков на результаты контент-анализа. Если контент-анализ не свободен от ценностей, кодировщики имеют меньше шансов прийти к соглашению относительно распределения категорий. Теорема возможности, применимая к ситуации выбора, обусловленного ценностями, утверждает, что "для любого метода определения социальных выборов агрегированием образцов индивидуальных предпочтений, который удовлетворяет определенным естественным условиям, возможно найти образцы индивидуального предпочтения, исключающие линейное упорядочение" [22, с. 330].

ка. Коэффициент *S* чувствителен к используемым в кодировании категориям, но ничего не говорит о совокупности данных [30, с. 248]. В сравнении с α и π, *S* ближе к к, потому что также не принимает в расчет предположения об "истинном" распределении категорий. По сравнению с к, *S* уделяет меньше внимания индивидуальным особенностям и пристрастиям кодировщиков. Вместо этого в случае использования показателя *S* ожидаемое согласие между кодировщиками рассчитывается при предположении о равной вероятности применения определенной категории. Если имеются две категории и два кодировщика, тогда вероятность достижения ими согласия равна $2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$ (вероятность того, что два независимых события в двух симметрично расположенных клетках матрицы сопряжения наступят одновременно).

Объяснение предположения о равной возможности применения определенных категорий может быть лучше понято при взгляде на контекст, в котором этот критерий надежности был первоначально предложен. Беннет и соавторы намеревались сравнить ответы, данные респондентами на вопрос по одной и той же теме, который был задан в формате, с одной стороны, закрытых вопросов в массовом опросе, а с другой стороны, полуструктурированных вопросов глубинного качественного интервью [26, с. 305]. "Истинного значения" (истинного распределения ответов) не существует в этих условиях, потому что каждый метод сбора данных имеет свои преимущества и недостатки. Возможно, похожая ситуация характеризует большинство непарадигмальных наук с отсутствием прочного консенсуса среди ученых, работающих в этих областях. В определенном смысле к их числу можно отнести социальные науки. Это направление рассуждений может объяснить высокую популярность критерия *S*: по подсчетам Криппендорфа, начиная с середины 1950-х годов, этот показатель изобретался заново в различных формах по меньшей мере пять раз [30, с. 245; 29, с. 80], в том числе в виде критерия *I,* обсуждаемого ниже.

С технической точки зрения, значение *S* зависит от *k,* т.е. числа категорий (A_o означает наблюдаемое согласие): $S = \frac{k}{k-1}\left(A_0 - \frac{1}{k}\right)$ [26, с. 307]. По этой причине он вызывает следующую критику: величина *S* имеет тенденцию к росту по мере увеличения количества неиспользуемых категорий, которые автор инструмента придумал, но редко использует при анализе данных [29, с. 80]. Тем не менее, данное направление критики предполагает, что "идеальная" книга кодов (та, которая исчерпывающе соответствует совокупности данных) существует. Подобная критика также не учитывает, что α не может быть подходящим стандартом для оценки "инфляции" *S*, поскольку α имеет тенденцию к дефляции, особенно при преобладании в кодировании одной категории.

Исходная формула для расчета *S* относится к случаю двух кодировщиков и номинальной категории (переменной) с двумя признаками (например, да/нет). *S* может быть рассчитан в общем виде как сумма расстояний от ситуации согласия между любым числом кодировщиков, которое может быть достигнуто при предположении равной вероятности категорий и их признаков: $S = 1 - \frac{1 - \sum_{i=1}^{N} \frac{A_{o_i} - A_{e_i}}{1 - A_{e_i}}}{N}$, где A_{o_i} означает наблюдаемое согласие между парой кодировщиков при применении категории *i*, A_{e_i} – к ожидаемому согласию между ними. $A_{e_i} = \frac{1}{2}$ в случае номинальной категории с двумя признаками, которая служит упрощению общей формулы: $S = 1 - \frac{2\sum_{i=1}^{N} A_{o_i}}{N} = 1 - \frac{2(b+c)}{N}$, где $(b + c)$ означает случаи несогласия между кодировщиками (значения клеток таблицы сопряженности, лежащих вне диагонали).

Моделирование использования набора данных, содержащего 12 единиц, одной номинальной категории с двумя признаками и нескольких кодировщиков (2, 3 и 4) показывает, что α имеет тенденцию к искусственному занижению, или дефляции (рис. 1). Независимо от количества единиц, по поводу которых наблюдается несогласие (в рассматриваемом случае оно варьирует от 0 до 12), α не превышает 0, что указывает на уровень согласия более низкий, чем тот, который мог быть достигнут чисто случайно. *S*, в свою очередь, варьирует от +1 до –1 (когда работают два кодировщика), хорошо реагируя на изменения количества кодировщиков и количество случаев, по поводу которых мнения кодировщиков расходятся. На значения *S* влияет только число случаев, в отношении которых мнения кодировщиков расходятся.

Показатель I_r Перроя и Лея. С количественной точки зрения, I_r представляет собой квадратный корень от *S* [32, с. 526]. Перро и Лей используют алгоритм для расчета I_r, который, однако,

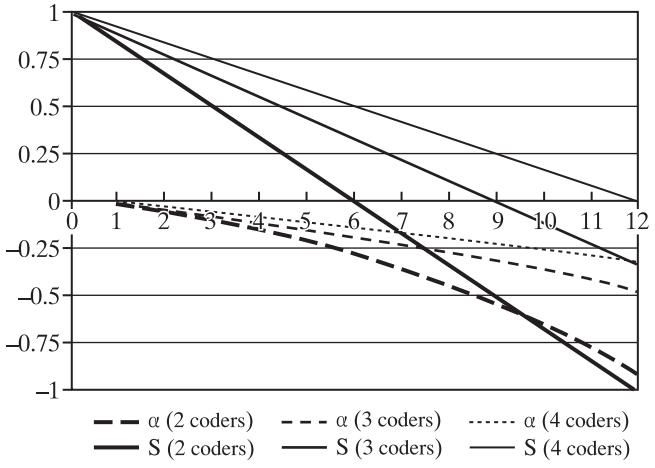

Рис. 1. Значения Alpha и *S* в зависимости от количества фрагментов, в отношении которых отсутствует согласие между кодировщиками; случаи 2-х, 3-х и 4-х кодировщиков; 12 кодируемых фрагментов.

значительно отличается от формулы Беннета и соавторов. “Надежность здесь может пониматься как процент от общих ответов (наблюдений), которые типичный кодировщик мог бы получить с учетом природы наблюдений, схемы кодирования, определений категорий, полученных инструкций и своих способностей” [40, с. 140].

Описанный таким образом I_r позволяет учесть и объективные, и субъективные факторы. Авторы помещают I_r на полпути между двумя точками отсчета: параметрами исследуемой совокупности и индивидуальными характеристиками кодировщиков (т.е. где-то между α и к). Согласно Перро и Лею, эти предположения соответствуют ситуации в маркетинге, где, с одной стороны, распределение ответов неизвестно заранее, и, с другой стороны, кодировщики стремятся получить объективную картину за счет минимизации влияния индивидуальных предпочтений и предвзятостей.

Показатель *r* Пирсона. Использование коэффициента Пирсона (*r*) в контент-анализе связано с рядом противоречий. Криппендорф осуждает идею коррелирования результатов контент-анализа, полученных несколькими кодировщиками [30, с. 245]. Поскольку он оспаривает то, что использование коэффициентов корреляции (α Кронбаха как главный пример) весьма спорно: качество работы одного кодировщика не должно оцениваться посредством сравнения с результатами работы другого. Когда кодировщики делают свою работу независимо друг от друга, решение одного не может “определять” выбор другого каким-либо значимым образом.

Нет необходимости, однако, предполагать причинную связь каждый раз, когда проводится корреляционный анализ. Существование зависимости между переменными – это необходимое, но недостаточное условие для предположения причинной связи между ними [27, с. 23]. Кроме того, корреляция может осуществляться не только между результатами работы кодировщиков.

Другие авторы не исключают использование корреляции в контент-анализе. В полезном сравнении *R*- и *Q*-режимов в дизайне исследований Дикстра и Эйнаттен [28] отмечают, что корреляции обычно используются в проектах *R*-режима в целях анализа сходства переменных в исследуемых объектах. В контент-анализе текст или визуальное изображение – единица исследования, тогда как переменные относятся к разным параметрам текста/изображения (первоисточник, жанр, дата и т.д.), вплоть до имен кодировщиков. *Q*-режим, который направлен на анализ сходства между объектами, также не исключает, по мнению указанных авторов, корреляционного анализа. Например, межклассовая корреляция может быть осуществлена, когда несколько кодировщиков используют большое число категорий в *Q*-режиме [28, с. 764]. Уорнер [45, с. 831–832] поддерживает такое мнение, добавляя, что корреляционный анализ применим только к количественным категориям (измеренным на порядковом, интервальном и относительном уровнях).

Олейник [36] предлагает вычисление корреляций между результатами контент-анализа и избранными количественными параметрами исследуемых текстов (например, показателями совместной встречаемости слов). Этот подход помогает избежать ряда проблем, на которые обращает внимание Криппендорф. Он также способствует нахождаению баланса между отсылками к “истинной оценке” (неотъемлемые характеристики текста) и интересами, намерениями и способностями кодировщиков. Следующая часть содержит аргументы для подтверждения уместности и разумности этой формы корреляционного анализа в контексте контент-анализа текстов.

ОСОБЕННОСТИ КОНТЕНТ-АНАЛИЗА ТЕКСТОВ

В отличие от контент-анализа открытых вопросов и других естественных единиц, выбор единицы кодирования (*unitizing*) является проблематичным в контент-анализе текстов и изображений [23, с. 580; 36, с. 872]. Выбор единицы кодирования предполагает выявление фрагментов текста, изображения, видеозаписи, содержащих

значимую для ответа на вопрос исследования информацию [30, с. 219]. “Для достижения совершенной надежности единицы, которые определяют различные наблюдатели, должны занимать одинаковое место в континууме и быть соотнесены с идентичными категориями” [31, с. 792]. Насколько нам известно, существующие компьютерные программы не предлагают опцию расчета показателей надежности при выборе единицы кодирования, что приводит к занижению остальных показателей надежности. Надежность выбора единицы кодирования, как правило, оказывается “спутывающей” *(confounding*) переменной при оценке надежности кодирования и выявлении факторов последнего, таких как ясность инструкций для кодировщиков.

Помимо технических особенностей, контент-анализ текстов также имеет более существенные характеристики. Текст как средство коммуникации может иметь несколько значений. При написании текста его автор передает сообщение. При чтении текста его читатели – во множественном числе, так как тексты обычно имеют несколько читателей – могут попытаться “расшифровать” сообщение автора.

Отметим, что в психологии активно прорабатываются подходы, связанные с проблематикой чтения: они основаны на рассмотрении коммуникативных навыков и навыков переработки информации, психологических процессов, связанных с целеполаганием и планированием, концентрацией внимания, регулированием когнитивных процессов при изменяющихся условиях, с самовосприятием и самокорректировкой и т.д. Эти подходы развиваются в сфере педагогики, образования для взрослых, библиотечного дела, концепциях эффективного, рационального, быстрого чтения.

Так, еще в 1920–1930-е годы развивалась библиопсихология, изучавшая чтение как сложный синтез индивидуально-психологических особенностей читателя и его социального опыта, а также социальных условий в целом, позиций социальных групп, которые стоят за автором текста и читателем, эмоциональной составляющей и др. [3, 15]. Это направление получило свое развитие в психологии и социологии чтения. Так, разделы, связанные с факторами понимания и интерпретации текстов, вошли в учебные курсы по психологии чтения (см., напр.: [10, 19] и др.).

Психологические теории чтения и соответствующие эмпирические исследования касаются различных вопросов восприятия, понимания и интерпретации текста. Среди них – взаимосвязь личностной и социальной составляющих, субъективная динамическая модель в имплицитных теориях чтения [8]; герменевтические проблемы множественности понимания и интерпретации текста, предопределенной его многоуровневостью и многообразием структур, смыслов, значений и функций, постмодернистской свободой интерпретаций (см.: [14]); рассмотрение техник понимания текстов в рамках риторико-герменевтического подхода к тексту и роли активного читателя, или развитой языковой личности [2]; анализ полипарадигмальности чтения и его экзистенциально-коммуникативных основ [17, 18] и др. Психологические вопросы академического чтения рассматривались и в контексте стратегиального подхода к восприятию научного текста [16, 13]. Этот краткий обзор российских исследовательских направлений в психологии и социологии чтения дает основания утверждать, что исследования понимания и интерпретации текстов в психологии могут предоставить достаточно разработанный базис для методологии контент-анализа. В свою очередь, независимые методологические разработки в области контент-анализа также представляют интерес для дальнейшего развития психологии чтения.

Читатели могут также привнести новые прочтения текста, открывая в нем новые идеи в зависимости от багажа собственных знаний, интересов, ценностей и способностей.

В связи с этим Норрис и Филипс различают *понимание* и *интерпретацию* (рис. 2). “При понимании человек стремится понять смысл текста, который был заложен автором в его версии пости-

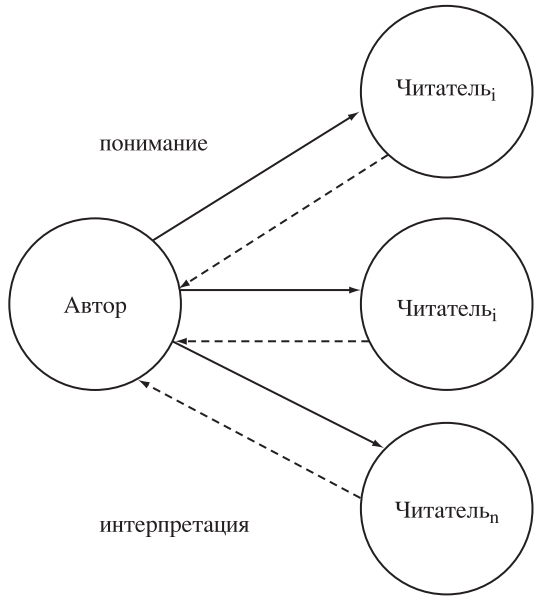

Рис. 2. Контент-анализ текста как совокупности взаимодействий между автором и читателем.

жения мира. При интерпретации человек раскрывает значение, которое не обязательно заложенно в самом объекте" [35, с. 402].

Анализ текста читателями каждый раз предполагает акцент либо на интерпретации, либо на понимании. Кроме того, различаются и форматы читаемых ими текстов. В семиотике различают два текстовых формата: *стилистические* и *риторические* [33, с. 45–51]. Риторические тексты (роман, поэма, дневник или сочинение) имеют рыхлую структуру. Метафоры и аналогии изобилуют в таких текстах. Контент-анализ риторических текстов обусловливает приоритетность интерпретации, ведь они направлены на провоцирование свободных ассоциаций и творческого мышления.

В отличие от них, стилистические тексты (научная статья, учебник или научное сообщение) имеют четкую, часто жесткую структуру. Аргументы в стилистических текстах должны отвечать высоким логическим стандартам: полноты и взаимной исключительности категорий, транзитивности и последовательности в их ранжировании и др. Понимание, как представляется, больше подходит для контент-анализа стилистических текстов ввиду их ориентации на передачу сообщения наиболее ясным и недвусмысленным способом. Интерпретация стилистических текстов, как и понимание риторических текстов, не исключаются, но представляются особыми случаями.

Таким образом, контент-анализ текстов должен, с одной стороны, быть помещен в контекст взаимодействия между автором и читателем, а с другой – принимать во внимание характеристики изучаемых текстов. Криппендорф признает, что текст может быть прочитан (и, следовательно, его содержание проанализировано) несколькими способами. В частности, он рассматривает три альтернативных допущения, лежащих в основе контент-анализа: содержание (1) заложено в тексте, (2) производно от источника текста и (3) возникает в процессе анализа исследователем текста в конкретном контексте [30, с. 19]. Затем он утверждает, что только третье предположение справедливо, однако: "одной из основных эпистемологических ошибок следует признать то, что тексты имеют присущие им значения или говорят сами за себя" [31, с. 789]. Преднамеренное ограничение контент-анализа интерпретацией делает кодирование менее неопределенным, но это достигается ценой исключения понимания из задач контент-анализа. Это также создает эпистемологические несоответствия в трактовке различных мер надежности: часто критикуемые как "слишком субъективные" κ и *S* лучше соответствуют духу интерпретации, чем более "позитивистские" π и α.

Цель настоящего исследования, сформулированная во введении, может быть теперь конкретизирована. Настоящая статья направлена на изучение вопроса о том, что является наиболее подходящим показателем надежности в зависимости от формата текста (стилистический или риторический) и типа чтения (понимание или интерпретация). Для этого было бы достаточно сосредоточиться на мерах надежности, применимых к (1) номинальным дихотомическим категориям (они чаще всего используются в контент-анализе текстов), (2) случаю нескольких кодировщиков (потому что текст потенциально имеет неограниченное число читателей)[3] и (3) различным точкам отсчета (замысел автора или точка зрения читателя).

Контент-анализ стилистических текстов включает в себя выбор сообщения автора в качестве точки отсчета, что предполагает предпочтение π и α, а не κ и *S*. Замысел автора задает параметры "истинного значения" (распределение категорий, соответствующее намерениям автора). Кодировщики достигают согласия, если они правильно понимают, что хотел сказать автор при помощи книги кодов и ее последовательного применения в процессе качественного кодирования.

При отправке "сообщения" в формате стилистического текста выбор слов определяет для автора "область возможного". При этом предположении становится оправданным "исследование ряда вещей, которые произносящий способен осуществить с помощью конкретных слов и предложений" [44, с. 3]. Меры совместной встречаемости слов (например, коэффициент косинуса сходства между текстами, см. [41, с. 203]) дают исследователю примерное представление о масштабах пересечений областей возможного, задаваемых различными текстами. Другими словами, показатели совместной встречаемости слов помогают определить относительные параметры "истинного значения". Существование корреляции между распределением качественных кодов в текстах (измеренной, например, с помощью коэффициентов сходства *косинус* или коэффициента сходства Жаккарда) и совместной встречаемости слов в

[3] Этот аргумент не стоит смешивать с другим, более "позитивистским", сформулированным Криппендорфом: "Мы должны оценивать распределение категорий в генеральной совокупности на основе мнений как можно большего числа кодировщиков (минимимум двух), делая стандартное предположение об усреднении различий между наблюдателями" [30, с. 249].

Таблица 1. Показатели надежности в зависимости от целей чтения и типа текста

Формат текста / Тип чтения	Стилистический	Риторический
Понимание	π, α, r (между результатами анализа совместной встречаемости слов и результатами качественного кодирования)	I_r
Интерпретация	I_r	S, к, r (между результатами кодирования с использованием основанного на замещении словами и результатами качественного кодирования)

этих же текстах указывает на надежный и достоверный характер качественного кодирования стилистических текстов [36, с. 868–871]. Понятие достоверности здесь используется в смысле соответствия между тем, что автор намерен донести до читателя, и тем, что читатель узнаёт из текста.

В свою очередь, точка зрения читателя преобладает в контент-анализе риторических текстов, что объясняет противоположный порядок предпочтений: к и S, а не π и α. Можно утверждать, что предположение о взаимозаменяемости кодировщиков не выполняется при контент-анализе риторических текстов. Каждый новый читатель привносит новые перспективы в прочтение текста. Следовательно, определение “истинного значения” становится менее актуальным, чем согласие между кодировщиками относительно того, что кажется важным для них. Кодировщики могут создавать надежные схемы кодирования, даже если книга их кодов не совпадает с категориями, которые автор имел в виду при написании текста. По этой причине показатели совместной встречаемости слов менее полезны для оценки надежности кодирования риторических текстов. Вместо того чтобы обращать внимание на совместную встречаемость слов, контент-аналитик может решить классифицировать их и создать основанный на замещении категорий словарь, структура которого соответствует книге кодов для качественного кодирования. Существование корреляции между распределением качественных кодов и распределением категорий, основанных на замещении словаря, в текстах указывает на надежный и достоверный характер качественного кодирования риторических текстов [36, с. 868–871]. Понятие достоверности в данном контексте используется в смысле соответствия между тем, что читатель хочет узнать из текста, и тем, что он действительно получает из него.

Эта цепь рассуждений позволяет получить предварительную классификацию показателей надежности в зависимости от формата текста и типа чтения (табл. 1). Достоверность контент-анализа, нацеленного на надлежащее понимание стилистических текстов, лучше измерять с помощью π, α и r (между показателями совместной встречаемости слов и результатами качественного кодирования). Надежность контент-анализа, нацеленного на интерпретацию риторических текстов, лучше измерять с помощью S, к, и r (между результатами контент-анализа с использованием основанного на замещении словаря и качественного кодирования). Наконец, в двух частных случаях (интерпретация стилистических текстов и понимание риторических текстов) I_r представляется наиболее подходящей мерой из-за ее промежуточного положения между параметрами совокупности и индивидуальными особенностями кодировщиков.

ИСТОЧНИКИ ДАННЫХ

Исчерпывающий ответ на сформулированный выше вопрос исследования требует контент-анализа двух совокупностей текстов – стилистических и риторических. На данном этапе мы можем обобщить результаты контент-анализа выборки стилистических текстов (научных статей, рецензий на книги и эссе) с использованием специализированной компьютерной программы *QDA Miner* с модулем *WordStat*. Программа исследования позволяет проверить две конкретные *гипотезы*, вытекающие из сформулированного выше вопроса. *Во-первых*, коэффициент корреляции Пирсона r между распределением качественных кодов в текстах и показателями совместной встречаемости слов в этих же текстах может быть использован для оценки надежности и достоверности контент-анализа стилистических текстов. В частности, этот коэффициент корреляции позволяет оценить степень совпадения результатов качественного кодирования с “истинным значением” текста (распределением категорий, соответствующих замыслу автора). *Во-вторых*, по сравнению с α,

показатель *S*, как правило, более тесно связан с распределением качественных кодов ввиду предполагаемой релевантности этого показателя для интерпретации текста. К сожалению, программа *QDA Miner* не может быть использована для расчета ни к, ни I_r, которые были бы еще более уместны в данном контексте.

В рамках научно-исследовательского проекта по академическому чтению четверо его участников, авторы настоящей статьи, проанализировали содержание выборки собственных научных публикаций. Следует отметить, что у участников проекта не было опыта предыдущего сотрудничества в научной сфере, и они работали в прошлом независимо друг от друга. В общей сложности в выборку были включены 57 текстов, опубликованных в период между 1999 и 2011 годами (20 написанных первым соавтором, социологом и экономистом, 20 – вторым соавтором, также социологом и экономистом, и 17 – третьим, социологом)[4]. Задача участников проекта, являвшихся одновременно кодировщиками, заключалась скорее в понимании текстов, чем в их интерпретации. Включение в выборку нескольких научных эссе (в отличие от научных статей), однако, означало, что элементов интерпретации избежать не удалось.

Контент-анализ проводился в несколько этапов. На первом этапе кодировщики читали тексты и разрабатывали собственные книги кодов для качественного кодирования всех текстов независимо друг от друга. Всем были предоставлены единые письменные инструкции по использованию компьютерной программы для качественного и количественного контент-анализа *QDA Miner* версии 4.0.4 с модулем *Wordstat* версии 6.1.4. Надежность и достоверность качественного кодирования оценивалась с помощью коэффициента *r*, рассчитанного между, с одной стороны, показателями распределения качественных кодов в текстах, и, с другой стороны, показателями совместной встречамости слов и распределения категорий, определенных с помощью словарей, основанных на замещении, в этих же текстах. Мы продолжали наше исследование только после достижения умеренно сильного уровня корреляции в каждом случае.

На втором этапе кодировщики разработали три общие книги кодов (одна адаптирована к текстам первого соавтора, другая – второго и третья – к текстам третьего), а также три общих словаря, основанных на замещении. Они содержали набор категорий, который затем был применен при повторном чтении кодировщиками той же выборки текстов независимо друг от друга. На этот раз в дополнение к коэффициентам корреляции *r* (так же, как на первом этапе), был рассчитан и коэффициент α. Исследование было продолжено только после достижения умеренно сильного уровня корреляции в каждом случае и значений коэффициента согласия между кодировщиками, превышающими 0.5 для каждой их пары[5].

Наконец, на третьем этапе были объединены три книги кодов. То же самое относится и к общему словарю, основанному на замещении. Кодировщики использовали общую книгу кодов при кодировании всех текстов. Другими словами, они попытались определить фрагменты, относящиеся к категориям, специфичным для конкретного автора, не только в текстах этого автора, но и в текстах, написанных другими авторами. Мы рассчитали коэффициенты корреляции *r* и коэффициенты согласия между кодировщиками α и *S*. Далее будут представлены результаты контент-анализа на втором и третьем этапах. Результаты первого этапа анализировались ранее [11]. Применение общей книги кодов в ее двух вариантах позволило нам оценить критерии надежности, а также достоверность понимания (на втором этапе) и интерпретации текста (на третьем этапе).

ЭМПИРИЧЕСКАЯ ПРОВЕРКА: ОЦЕНКА НАДЕЖНОСТИ КОНТЕНТ-АНАЛИЗА НАУЧНЫХ ТЕКСТОВ

Результаты корреляционного анализа показывают, что четыре кодировщика были последовательны в применении обеих версий книги кодов. Это выразилось в том, что распределение качественных кодов коррелирует с совместной встречаемостью слов (табл. 2). Полученный результат дает основание утверждать, что распределение категорий соответствовало замыслам авторов (отраженным в выборе слов и их сочетаний). В данном случае значения *r* оказались выше, чем в исследовании, проведенном первым соавтором ранее с использованием аналогичного подхода. Тогда было проанализировано содержание текстов качественных частично структурированных

[4] Четвертый участник проекта и соавтор статьи, недавняя выпускница вуза, пока еще не написала достаточного числа статей. Она выполняла роль “идеального читателя”, чье восприятие включенных в выборку текстов не было искажено авторством некоторых из них.

[5] При оценке достигнутого уровня согласия между кодировщиками следует учесть, что он отражает как надежность кодирования, так и надежность в выборе единиц кодирования.

Таблица 2. Значения коэффициентов корреляции *r* между распределением качественных кодов в текстах, совместной встречаемостью слов и распределением категорий, выявленных с помощью основанного на замещении словаря в тех же текстах, 4 кодировщика, $N = 57$

		Второй этап			Третий этап		
		Качественное кодирование	Основанный на замещении словарь	Совместная встречаемость слов	Качественное кодирование	Основанный на замещении словарь	Совместная встречаемость слов
Второй этап	Качеств-ое кодиров-ие	1	.747**	.674**	–	–	–
	Основ-ный на замещении словарь	.747**	1	.707**	–	–	–
	Совм-ая встреч-сть слов	.674**	.707**	1	–	–	–
Третий этап	Качеств-ое кодир-ние	–	–	–	1	.902**	.684**
	Основ-ный на замещении словарь	–	–	–	.902**	1	.707**
	Совместная встречаемость слов	–	–	–	.684**	.707**	1

** – Корреляция статистически значима на уровне значимости 0.001 (двухсторонняя значимость). Уровень статистической значимости приводится исключительно в целях сравнения, учитывая неслучайный характер выборки.

интервью – они содержали больше риторических элементов, чем научные тексты – и было обнаружено, что коэффициент корреляции *r* между распределением качественных кодов и совместной встречаемостью слов варьировал в пределах от 0,295 до 0,483 [36, с. 869].

Усиление корреляции между распределением качественных кодов и распределением категорий, определенных с помощью словаря, основанного на замещении, на третьем этапе нашего исследования может свидетельствовать о более сильном акценте на интерпретации[6].

Мы сравнили значения коэффициента корреляции *r* со значениями α и *S* для предоставления дополнительных доказательств в поддержку нашего мнения о необходимости добавления его к списку показателей надежности контент-анализа текстов (табл. 3).

На этот раз анализировались корреляции между распределениями качественных кодов, произведенных каждым кодировщиком. Определялось, насколько они схожи, с помощью расчета коэффициентов подобия Жаккарда между конкретным текстом и центроидом (текстом, занимающим центральное положение на плоскости двухмерного шкалирования по критерию совместной встречаемости кодов). Код использовался в качестве единицы анализа при расчете корреляции на третьем этапе. Это соответствует логике *Q*-режима в статистическом моделировании, в отличие от *R*-режима, для которого текст является более естественной единицей анализа. Общая книга кодов содержит 37 позиций[7].

На втором этапе применение кода в качестве единицы анализа при расчете корреляции, по нашему мнению, не имело бы особого смысла, так как в результате этой операции были бы получены три полностью независимых кластера кодов. Поэтому на этом этапе в корреляционном анализе как его единицу мы использовали текст с целью выяснить, насколько сходными (с точки зрения коэффициента подобия косинус) являются распределения случаев (текстов) по критерию совместной встречаемости кодов. Могло показаться, что это снизило степень методологической

[6] Словарь, основанный на замещении, на третьем этапе был подвергнут лишь незначительной корректировке.

[7] Книга кодов для контент-анализа текстов первого соавтора статьи включает в себя 13 кодов, книга кодов для контент-анализа второго соавтора – 15 кодов, третьего соавтора – 9 кодов.

Таблица 3. Значения α Криппендорфа, *S* Беннета, коэффициентов корреляции Пирсона *r* между распределением качественных кодов

Показатель / Кодировщики	Второй этап			Третий этап			
	α (*N*=37)	*S* (*N*=37)	*r* (*N*=57)	α (*N*=37)	*S* (*N*=37)	*r* (*N*=37)	*r* (*N*=57)
A+S	0.575	0.820	0.908	0.412	0.640	0.818	0.774
A+I	0.535	0.802	0.949	0.436	0.679	0.934	0.909
A+T	0.555	0.812	0.869	0.434	0.698	0.957	0.674
S+I	0.544	0.813	0.833	0.423	0.653	0.848	0.656
S+T	0.519	0.811	0.755	0.404	0.666	0.873	0.534
I+T	0.496	0.797	0.831	0.399	0.687	0.937	0.721
Среднее для пар кодировщиков	*0.537*	*0.809*	*0.858*	*0.418*	*0.671*	*0.895*	*0.711*
A+S+I	0.465	0.590	–	0.289	0.365	–	–
A+S+T	0.465	0.590	–	0.286	0.375	–	–
S+T+I	0.432	0.574	–	0.278	0.371	–	–
A+I+T	0.440	0.570	–	0.299	0.398	–	–
Среднее для триад кодировщиков	*0.4505*	*0.581*	–	*0.288*	*0.377*	–	–
A+S+T+I	0.367	0.420	–	0.208	0.245	–	–

Примечание. $N = 57$ соответствует числу закодированных текстов; $N = 37$ соответствует числу кодов, включенных в общую книгу кодов; *A, S, I* и *T* обозначают имена кодировщиков (авторов данной статьи).

"чистоты". Однако обе версии корреляционного анализа на третьем этапе (одна с помощью кода в качестве единицы анализа, а другая – текста как единицы анализа) показали сходящиеся, а не расходящиеся результаты (подробнее об этом см. [37]).

Другое ограничение относится к применимости корреляционного анализа только к парам кодировщиков. Двухмерные корреляции не могут быть использованы непосредственно для оценки степени согласия между тремя и более кодировщиками. В случае триады, квартета и так далее сравнения могут быть осуществлены лишь попарно, что объясняет, почему некоторые клетки таблицы 3 остались пустыми.

Можно видеть, что α оказалась умеренно связанной с *r* на втором этапе: коэффициент корреляции между ними был равен 0.503 ($N = 6$ пар)[8]. Отношения между показателями надежности на третьем этапе необходимо обсудить более подробно. По-прежнему α связан с *r*. Коэффициент корреляции Пирсона между ними составил 0.474 при использовании текста в качестве единицы анализа и 0.329 при использовании кода, в качестве единицы анализа (N=6 пар в обоих случаях). Таким образом, наша *первая гипотеза* (о том, что коэффициент корреляции Пирсона *r* между распределением качественных кодов в текстах и показателями совместной встречаемости слов может быть использован для оценки надежности и достоверности контент-анализа стилистических текстов) может быть предварительно принята: исследованный коэффициент корреляции не противоречит другим показателям надежности, а именно α.

Еще более существенно то, что *S* оказался почти идеально связан с *r*: коэффициент корреляции Пирсона между ними составил 0.985 при использовании кода в качестве единицы анализа. Этот факт позволяет предположить, что *S* и α действительно отражают различные аспекты согласия между кодировщиками. Показатель *S* более тесно связан с распределением качественных кодов, что подтверждает его наибольшую релевантность при интерпретации текста. Как уже отмечалось, в отличие от второго этапа, программа третьего этапа предполагала больший акцент на интерпретации, чем на понимании текста. На втором этапе *S* не был связан с *r* (коэффициент корреляции между ними равен -0.013). Таким образом, наша *вторая гипотеза* (о том, что показатель *S* более тесно связан с распределением качественных кодов и релевантен для интерпретации текста) может быть предварительно принята.

Как показано в таблице 3, значение α зависит от числа кодировщиков. Среднее значение этого

[8] Распределение показателей надежности было проинспектировано визуально перед проведением корреляционного анализа. Выяснилось, что это распределение близко по своим параметрам к нормальному.

показателя для пар кодировщиков было выше, чем среднее значение для триад. А среднее значение для триад было выше, чем α, рассчитанное для случая всех четырех кодировщиков. Этот факт опровергает предположение о взаимозаменяемости кодировщиков в контент-анализе текстов, в особенности для риторических текстов при задачах интерпретации (различия в средних значениях были особенно значимы на третьем этапе).

ЗАКЛЮЧЕНИЕ

Ограниченный характер нашей выборки не позволяет делать далеко идущие выводы. Тем не менее, полученные результаты скорее говорят в пользу, чем против предложения использовать коэффициенты корреляции для оценки контент-анализа текстов. Они дополняют другие показатели надежности, такие как коэффициенты α Криппендорфа, а также *S* Беннета и соавторов. Результаты нашего исследования также говорят о том, что выбор показателя надежности зависит от формата текста (стилистический или риторический) и типа чтения (понимание или интерпретация). В частности, *S* и κ как показатели надежности, мы считаем, больше всего подходят для перспективного прочтения риторических текстов (интерпретации).

В качестве наиболее интересных направлений для дальнейших исследований нам представляются следующие. *Во-первых*, более тщательная проверка гипотезы о зависимости выбора показателя надежности от формата текста и типа чтения. Это предполагает сравнение результатов *интерпретации* идеал-типичных риторических текстов (стихи, романы) с результатами *понимания* идеал-типичных стилистических текстов (научные статьи в парадигмальных науках).

Во-вторых, на методическом уровне требуется совершенствование методов расчета коэффициента *S* и его вариантов. Криппендорф [30, с. 223–249] обсуждает, по крайней мере, четыре метода расчета α и, соответственно, его четыре интерпретации. Арстейн и Поэзио [23, с. 565–566] используют сумму квадратов разностей от среднего для расчета α. Можно утверждать, что подобный подход может быть применен к случаю *S*. Это будет способствовать нахождению ответа на наиболее распространенную критику *S* касательно зависимости этого показателя от количества неиспользованных кодов.

В-третьих, необходимо рассмотреть случаи, когда показатели надежности, адаптированные к перспективному чтению (интерпретации), могут иметь новые, иногда неожиданные области применения. Они полезны в любой ситуации, в которой важно субъективное восприятие оценивающего. Формулировки решений, принятых коллегиями "судей", относятся к одной из таких областей. Если есть разногласия между тремя и более членами коллегии, то они часто предоставляют отдельные обоснования своей позиции. Анализ содержания таких документов с помощью π или α не имеет смысла: как и читатели риторических текстов, "судьи" по определению не являются взаимозаменяемыми. Таким образом, *S* и коэффициенты корреляции могут оказаться лучшей мерой для измерения их согласия. Термин "судья" употребляется в контент-анализе в переносном смысле как синоним кодировщика. Такими "судьями" могут быть члены команды, проводящей психологические эксперименты, анализирующие данные социологических интервью и т.д.

В-четвертых, в отличие от контент-анализа риторических текстов, изучение стилистических текстов легче поддается компьютеризации, как это показано в нашей работе (в частности, когда используются показатели совместной встречаемости слов). Это открывает путь для добавления новых опций к электронным базам научных публикаций, таких как *Web of Science* или *eLibrary,* особенно при рассмотрении групп текстов, относящихся к одной области – психологии, социологии и др. Одно из направлений дальнейшего анализа может состоять в категоризации научных текстов с использованием не только довольно ограниченного числа ключевых слов, но и на основе показателей совместной встречаемости слов в опубликованных текстах.

СПИСОК ЛИТЕРАТУРЫ

1. *Абраменкова В.В.* Отражение картины мира современных российских детей в графических и вербальных образах // Вопросы психологии. 2007. № 6. С. 54–63.
2. *Бушев А.Б., Александрова С.К.* Техники интерпретации текста как способности развитой языковой личности // Гуманитарный вектор. 2011. № 4. С. 39–51.
3. *Вальдгард С.Л.* Очерки психологии чтения. СПб.: Российская национальная библиотека, 2010 [1931].
4. *Володарская Е.А., Логвинова Н.Ю.* "Родительская" и "студенческая" модели представлений о семейном воспитании // Психологический журнал. 2005. Т. 26. № 5. С. 26–34.
5. *Воловикова М.И., Соснина Л.М.* Этнокультурное исследование представлений о справедливости (на примере молдаван и живущих в Молдове цыган) // Вопросы психологии. 2001. № 2. С. 85–93.

6. *Гулевич О.А.* Социальные представления о преступлениях в межличностной и массовой коммуникации // Вопросы психологии. 2001. № 4. С. 53–67.
7. *Кирдина С.Г., Шаталова Т.Ю.* Возрастающая отдача в современной экономической литературе: контент-анализ российских и зарубежных источников / Феномен возрастающей отдачи в экономике и политике. СПб.: Алетейя, 2013. С. 18–54.
8. *Кондаков И.М., Ушнев С.В.* Экспериментальное исследование имплицитных теорий чтения // Вопросы психологии. 1994. № 6. С. 110–117.
9. *Лысенко Н.Е., Давыдов Д.М.* Оценка текстовых описаний сцен насилия в зависимости от психотизма и половых различий // Психологический журнал. 2011. Т. 32. № 3. С. 114–127.
10. *Миронова М.В.* Психология и социология чтения: Учебное пособие. Ульяновск, 2003.
11. *Олейник А.Н., Кирдина С.Г., Попова И.П., Шаталова Т.В.* Как ученые читают друг друга: основы теории академического чтения и ее эмпирическая проверка // Социологические исследования. 2013. № 8. С. 30–41.
12. *Павлова Н.Д.* Новые направления в психологии речи и психолингвистике // Психологический журнал. 2007. Т. 28 № 2. С. 19–30.
13. *Пранцова Г.В., Романичева Е.С.* Современные стратегии чтения и понимания текста. М.: Форум, 2013.
14. *Пузько В.И.* Способы понимания и интерпретации текста / Учебное пособие. Владивосток, 2010.
15. *Рубакин Н.А.* Библиологическая психология и литературоведение. Доклад обществу любителей российской словесности, читанный 11 мая 1927 г. в Москве // Вопросы психолингвистики. 2007. № 6. С. 189–206.
16. *Рубо И.Г.* Психологический анализ стратегий чтения научного текста: Дисс. ... канд. психол. наук. М., 1988.
17. *Стефановская Н.А.* Методологические проблемы эмпирических социологических исследований чтения // Аналитика культурологии. 2007. № 7. С. 280–288.
18. *Стефановская Н.А.* Экзистенциальные основы чтения. Тамбов: Издат. дом ТГУ им. Г.Р. Державина, 2008.
19. *Тараканов А.В.* Психология и социология чтения / Учебное пособие. Новосибирск, 2011.
20. *Ушакова Т.Н., Белова С.С., Валуева Е.А.* Лингвопсихологическое исследование вербальной семантики // Психологический журнал. 2010. Т. 31. № 6. С. 83–97.
21. *Allport G.W.* The use of personal documents in psychological science. New York: Social Science Research Council, 1942.
22. *Arrow K.J.* A Difficulty in the Concept of Social Welfare // The Journal of Political Economy. 1950. 58. Is. 4. P. 328–346.
23. *Artstein R., Poesio M.* Inter-Coder Agreement for Computational Linguistic // Computational linguistic. 2008. V. 34. Is. 4. P. 555–596.
24. *Baldwin A.L.* Personal structure analysis: A statistical method for investigating the single personality // Journal of Abnormal and Social Psychology. 1942. № 37. P. 163–183.
25. *Bales R.F.* Interaction process analysis. Reading, MA: Addison-Wesley, 1950.
26. *Bennett E., Alpert R., Goldstein A.C.* Communications through Limited-Response Questioning // Public Opinion Quarterly. 1954. V. 18. Is. 3. P. 303–308.
27. *Bryman A., Bell E., Teevan J.J.* Social Research Methods. 3rd edition. Oxford University Press, Don Mills, 2012.
28. *Dijkstra L., Eijnatten F.M. van.* Agreement and Consensus in a Q-mode Research Design: an Empirical Comparison of Measures, and an Application // Quality & Quantity. 2009. V. 43 Is. 5. P. 757–771.
29. *Hayes A.F., Krippendorff K.* Answering the Call for a Standard Reliability Measure for Coding Data // Communication Methods and Measures. 2007. № 1 (1). P. 77–89.
30. *Krippendorff K.* Content Analysis: An Introduction to Its Methodology. SAGE, Thousand Oaks, 2004.
31. *Krippendorff K.* Measuring the Reliability of Qualitative Text Analysis Data // Quality & Quantity. 2004. № 38 (6). P. 787–800.
32. *Leiva F.M., Ríos F.J.M., Martínez T.L.* Assessment of Interjudge Reliability in the Open-Ended Questions Coding Process // Quality & Quantity. 2006. № 40 (4). P. 519–537.
33. *Lotman Y.* Universe of the Mind: A Semiotic Theory of Culture. Indiana University Press, Bloomington, 1990.
34. *Neuendorf K.A.* The Content Analysis Guidebook. SAGE, Thousand Oaks, 2002.
35. *Norris S.P., Philips L.M.* The Relevance of a Reader's Knowledge within a Perspectival View of Reading // Journal of Reading Behavior. 1994. № 26 (4). P. 391–412.
36. *Oleinik A.* Mixing Quantitative and Qualitative Content Analysis: Triangulation at Work // Quality & Quantity. 2010. № 45 (4). P. 859–873.
37. *Oleinik A., Popova I., Kirdina S., Shatalova T.* On the choice of measures of reliability and validity in the content-analysis of texts // Quality & Quantity. 2014. № 48. P. 2703–2718.
38. *Osgood C.E.* Probing subjective culture: Part 1. Cross-linguistic tool-making // Journal of Communication. 1974. № 24 (1). P. 21–35.
39. *Osgood C.E.* Probing subjective culture: Part 2. Cross-cultural tool-using // Journal of Communication. 1974. № 24 (2). P. 82–100.

40. *Perreault W.D., Leigh L.E.* Reliability of Nominal Data Based on Qualitative Judgments // Journal of Marketing Research. 1989. № 26 (2). P. 135–148.

41. *Salton G., McGill M.* Introduction to Modern Information Retrieval. McGraw-Hill Book Co., New York, 1983.

42. *Scott W.A.* Reliability of Content Analysis: The Case of Nominal Scale Coding // Public Opinion Quarterly. 1955. № 19 (3). P. 321–325.

43. *Siegel S., Castellan N.J.* Nonparametric Statistics for the Behavioural Sciences. McGraw Hill, New York, 1988.

44. *Skinner Q.* Visions of Politics. Cambridge University Press, Cambridge, 2002.

45. *Warner R.M.* Applied Statistics. SAGE, Thousand Oaks, 2008.

46. *Weller S.C.* Cultural Consensus Theory: Applications and Frequently Asked Questions // Field Methods. 2007. № 19 (4). P. 339–368.

47. *White R.K.* Black Boy: A value analysis // Journal of Abnormal and Social Psychology. 1947. № 42. P. 440–461.

RELIABILITY AND VALIDITY IN TEXTS' CONTENT-ANALYSIS: THE CHOICE OF MEASURES

A. N. Oleinik*, I. P. Popova, S. G. Kirdina***, T. Y. Shatalova******

* *Sc.D. (ecomonic), PhD, leading research officer, Central Economic Mathematical Institute RAS, Memorial University, Canada;*

** *PhD, senior research officer, Institute of Sociology RAS, leading research officer, National Research University – Higher School of Economics, Moscow;*

*** *Sc.D. (sociology), head of department, Institute of Economics RAS, Moscow;*

**** *lecturer, Moscow State University named after M.V. Lomonosov, Moscow*

The paper discusses several reliability measures: Scott's pi, Krippendorff's alpha, free marginal adjustment (Bennett, Alpert and Goldstein's S), Cohen's kappa, and Perreault and Leigh's I and the assumptions on which they are based. It is suggested that correlation coefficients between, on one hand, the distribution of qualitative codes and, on the other hand, word co-occurrences and the distribution of the categories identified with the help of the dictionary based on substitution complement the other reliability measures. The paper shows that the choice of the reliability measure depends on the format of the text (stylistic versus rhetorical) and the type of reading (comprehension versus interpretation). Namely, Cohen's kappa and Bennett, Alpert and Goldstein's S emerge as reliability measures particularly suited for perspectival reading of rhetorical texts. Outcomes of the content analysis of 57 texts performed by four coders with the help of computer program QDA Miner inform the analysis.

Key words: content-analysis, reliability measures, correlation analysis, interpretation, stylistic texts, rhetorical texts, psychology of reading.