

*Е.В. Кремнёв*

## НЕКОТОРЫЕ ПРОБЛЕМЫ ПРИМЕНЕНИЯ КОНТЕНТ-АНАЛИЗА К КИТАЙСКИМ ПИСЬМЕННЫМ ТЕКСТАМ

*В статье рассматривается вопрос о необходимости специфических подходов к применению контент-анализа к китайским письменным текстам в связи с некоторыми особенностями китайского языка. Основными проблемами при использовании контент-анализа являются правильное выделение единиц анализа и единиц счета применительно к китайскому тексту, а также верная числовая интерпретация полученных результатов на основе стилистических особенностей китайского языка.*

**Ключевые слова:** контент-анализ; китайский текст; единица анализа; единица счета.

*E. V. Kremnyov*

## THE PROBLEMS OF APPLICATION OF CONTENT ANALYSIS TO THE CHINESE WRITTEN TEXTS

*The article describes the need for specific approaches to the application of content analysis to the Chinese written texts in connection with some special features of the Chinese Language. While using content analysis, the main problems are the correct selection of analysis units and count units as well as correct interpretation of results based on stylistic special features of the Chinese texts.*

**Key words:** content analysis; Chinese text; analysis unit; count unit.

Контент-анализ определяют как «формализованный метод изучения текстовой и графической информации, заключающийся в переводе изучаемой информации в количественные показатели и ее статистической обработке» [Дмитриев, 2005]. Кроме того, его также можно определить как «метод выявления и оценки специфических характеристик текстов и других носителей информации (видеозаписей, теле- и радиопередач, интервью, ответов на открытые вопросы и т. д.), в котором в соответствии с целями исследования выделяются определенные смысловые единицы содержания и формы информации, а затем производится систематический замер частоты и объема упоминаний этих единиц в определенной совокупности текстов или другой информации» [Психологический словарь, 2000]. Этот метод, также называемый «количественным анализом документов», крайне необходим, когда речь заходит об исследовании скрытых закономерностей. В таких случаях особый интерес для исследователя представляет информация, сознательно или подсознательно включенная автором в текст помимо общей коммуникативной цели или наряду с ней.

При значительном количестве проанализированных источников контент-анализ позволяет оперировать достаточно точными данными, выраженными в числовых показателях. Это делает его незаменимым при исследовании самых различных явлений в ряде гуманитарных наук. К примеру, в психологии контент-анализ позволяет выявлять отдельные психологические характеристики коммуникатора, аудитории, сообщения и их взаимосвязи. При этом, в сравнении с элементарным содержательным анализом, контент-анализ как научный метод психологии дает информацию, которая отвечает определенным критериям качества, предъявляемым к результатам исследования: надежности, объективности, валидности. Контент-анализ активно применяется в социологии и политологии для изучения таких социально-политических и экономических процессов, которые внешне представляются хаотическими, бессистемными и имеют множество скрытых факторов влияния.

Важность таких исследований, проводимых касательно внутренних процессов в КНР, очевидна. При всей открытости, которую сегодня позиционирует Китай, большая часть

важной информации остается завуалированной, скрытой. В таких исследованиях контент-анализ, по утверждению социолога А.Г. Здравомыслова, является «научно обоснованным методом чтения между строк», может быть применен в самых различных вариациях: как в соединении с качественным, так и в виде сугубо количественного анализа. Кроме того, следует отметить, что контент-анализ нередко применяется вкупе с другими методами исследования. В таких случаях либо он выступает в качестве проверочного метода, либо данные, полученные с его помощью, сравнивают с данными, полученными с использованием других методов. Заметную роль в повышении качества контент-анализа играет возможность использования методов многомерного статистического анализа данных. Особенно широко используется факторный анализ, способствующий выявлению скрытых факторов, определяющих содержание текстов.

Вместе с тем, в процессе применения контент-анализа к китайским текстам исследователь чаще всего рискует столкнуться с проблемами, игнорирование которых может привести как к искажению информации на стадии ее получения, так и к неверной интерпретации. Рассмотрим некоторые из таких проблем на примере количественного анализа, основанного на подсчете частности употребления единиц анализа, выраженных лексическими единицами, и их взаимной встречаемости. В таких случаях единицы анализа будут приравниваться к единицам счета. Поскольку каждый этап характеризуется своей проблемой, рассмотрим трудности в зависимости от процедур, при которых они возникают.

Количество и качественные характеристики процедур контент-анализа могут различаться, однако в подавляющем большинстве исследований при работе с вышеуказанной разновидностью контент-анализа к основным процедурам относят следующие: выявление смысловых единиц контент-анализа, выделение единиц счета, собственно процедуру подсчета и интерпретацию полученных числовых данных.

На этапе выявления единиц анализа и единиц счета встает серьезная проблема определения формы самой единицы. Данная трудность заложена в структурных различиях ки-

тайского и европейского текста. В отличие от текстов, написанных на европейских языках, китайский имеет качественно иную структуру. Основной единицей текстового материала является иероглифический знак. В связи с этим исследователю приходится выбирать между подходами, основанными на различных принципах к выделению базовых единиц, а выбор этот всегда продиктован самим текстом. Первый подход заключается в выделении в качестве единиц анализа монономов, то есть слов, представленных в тексте одним иероглифическим знаком. Очевидно, что такой знак может быть единицей анализа только тогда, когда того требуют поставленные исследователем задачи. Это касается работ, непосредственно изучающих знак, а также тех случаев, когда анализируемый текстовый массив представляет собой произведения, созданные целиком на вэньяне. Примером может служить отрезок из «Дао дэ цзин», в котором большая часть иероглифических знаков может стать единицами анализа и единицами счета:

«道，可道，非恒道。名，可名，非恒名。无名，天地之始；有名，万物之母。故常无欲，以观其妙；常有欲，以观其徼。此两者同出而异名，同谓之玄。玄之又玄，众妙之门»。

За единицу анализа здесь, в зависимости от задач исследования, могут быть взяты любые знаки (道, 名, 天, 地), являющиеся полнозначными словами, а единицами счета – как сами единицы анализа, если речь идет о подсчете частности встречаемости, так и другие характеристики текста (например, взаимная встречаемость знаков 道 и различных отрицаний: 非, 无 и т.д.). Таким образом, текст, в котором знак равняется слову, не вызывает затруднений с выбором единиц анализа и единиц счета.

Второй подход основан на выборе в качестве единиц контент-анализа как монономов, так и полиномов, то есть слов, записываемых несколькими иероглифами. Как правило, этот подход применяется в тех случаях, когда мы имеем дело с материалами, составленными на современном китайском языке, поскольку большая часть лексики в таких текстах представлена словами, записываемыми как одним, так и двумя и более иероглифами. Это, в первую очередь, затрудняет применение автоматизированных систем при последующем про-

цессе подсчета единиц, поскольку большинство из них построено на подсчете слов, разделенных пробелом, чем китайский текст также не обладает. Большинство исследователей при этом сначала используют дробление текста на слова, а уже затем приступают к подсчету. Следует отметить, однако, что по большей части такие методики не полностью удовлетворяют требованиям точности подсчета, поскольку дробление текста является весьма условным, что может привести к значительным отклонениям числовых закономерностей и, в итоге, ошибочным выводам, сделанным на их основе.

Решение этой проблемы, как нам представляется, лежит в двух противоположных направлениях. Первое направление заключается в полном подсчете всех единиц счета без использования автоматизированных средств. Такой процесс, хотя и является трудоемким и протяженным во времени, тем не менее дает достаточно точные результаты, поскольку учитывает все способы употребления лексических единиц. Вторым направлением может стать разработка и использование специальных программ типа CLAAS (Chinese Language Automatic Annotation System), позволяющих проводить предварительное реферирование текста с выбором ключевых слов, словосочетаний и предложений с заданными параметрами [Загибалов, 2005]. В этом случае программа позволяет проводить подсчет заданных единиц без дробления текста, увеличивая точность подсчета.

На последующем этапе при проведении структурного анализа текста и выявлении частотности упоминания того или иного термина исследователь сталкивается с еще одной проблемой, связанной со стилистическими особенностями китайских текстов. При проведении данной процедуры необходимо учитывать тот факт, что частотность употребления ключевых слов в китайском и русском текстах различны. Повторы, недопустимые в русском и большинстве европейских языках, в китайском – привычное явление.

В качестве примера можно рассмотреть отрезок текста из статьи на китайском языке, представляющий определение понятия «领导文化» в самом широком смысле и перевод этого отрезка:

«领导文化是一国社会政治文化的特殊分布, 是领导者群体或个体在领导实践中形成的, 对于领导活动的过程、本质、规律、规范、价值以及方式方法等各方面的综合反映形式, 是领导者开展领导活动和从事领导行为的内驱动力和精神导向» [Бай Вэйчунь, 2001].

*«Управленческая культура – это особая составляющая социально-политической культуры страны, формирующаяся в процессе коллективной или личной управленческой практики, комплексно отражающая процесс, сущность, законы, нормы, значение управленческой деятельности, ее формы и способы, и являющаяся внутренней стимулирующей силой и духовной направляющей деятельности управленцев».*

Из примера видно, что для адекватного перевода переводчику не требуется употреблять слова «управление» и «управленческий» такое же количество раз для передачи того же спектра значений, что дается в китайском определении.

Кроме того, китайский пример можно также сравнить с самостоятельным аналогичным определением, приведенным в статье русскоязычного автора:

*«Понятие “управленческая культура” интерпретируется нами как целостное единство управленческих знаний, чувств, ценностей, управленческих и организационных отношений на определенном этапе управленческой деятельности, обеспечивающее достижение поставленных целей. И в настоящее время управленческая культура существенно влияет на профессиональную деятельность руководителя, так как является важнейшим критерием, показателем профессионализма и компетентности современного руководителя» [Начкебия, 2008].*

В китайском определении слово «领导» применяется в нескольких значениях, в русском же определении эти значения передаются разнокоренными словами: «управленческий» и «руководитель», в связи с чем они становятся разными единицами счета и могут в зависимости от требований исследования войти в разные категории анализа.

Таким образом, анализ частотности возможен только при учете этого фактора. Это важно и на завершающем этапе контент-анализа: интерпретации полученных числовых дан-

ных, особенно если имеет место использование числовых результатов контент-анализа для сравнения частотности употребления тех или иных терминов в текстах на китайском языке с частностью в текстах на русском и европейских языках. Кроме того, даже при работе только с китайскими текстами установление верных числовых соотношений возможно только при условии выявления частности нескольких смысловых единиц для каждого текста. Это особенно важно в том случае, если контент-анализ служит вспомогательным методом для проверки данных, полученных иными методами. В этом случае адекватность оценки данных зависит от верного понимания специфики языка и правильной интерпретации числовых данных в рамках этой специфики.

Указанные проблемы далеко не исчерпывают тот ряд трудностей, что встают перед исследователем в работе над китайским текстом, они лишь сужены до анализа частотности употребления единиц анализа, приравненных к единицам счета. Проведение контент-анализа в различных его вариациях в отношении китайских текстов неизбежно будет ставить перед исследователем проблемы, сущность которых состоит в том, что изначально контент-анализ был разработан для текстов, написанных на европейских языках. Тем не менее это вовсе не означает, что данный метод в отношении китайского языка неприменим. Более того, контент-анализ при правильном подходе дает исчерпывающие и точные данные. Однако получение таковых от анализа китайских текстов возможно только в том случае, если применение метода проводится с учетом ря-

да специфических характеристик китайского языка. Изучение этих особенностей дает исследователю возможность избежать ошибок на каждом из этапов анализа и получения неверных результатов при проведении исследования с использованием контент-анализа.

#### Библиографический список

1. *Бай Вэйчунь*. Дандай чжунго линдао вэньхуа дэ цзегоусин тэчжэн луньси (Анализ структурных особенностей современной культуры госаппарата КНР) [Электронный ресурс] / Бай Вэйчунь, Чжан Шунь // Научный вестник Северовосточного университета. – 2001. – №6. – Режим доступа: <http://emuch.net/journal/article.php?id=CJFDTotalDBSS200106006>.
2. *Дмитриев, И.В.* Контент-анализ: сущность, задачи, процедуры [Электронный ресурс] / И.В. Дмитриев // ПСИ-ФАКТОР. – 2005. – Режим доступа: <http://www.psyfactor.org/lib/k-a.htm>.
3. *Загibalов, Т.Е.* Автоматический анализ текстов на китайском языке. Проблема выбора базовой единицы [Электронный ресурс] / Т.Е. Загibalов // Диалог: междунар. конф. по компьютерной лингвистике / Красноярский государственный университет». – 2005. – Режим доступа: <http://www.dialog-21.ru/Archive/2005/Zagibalov%20T/Zagibalov.htm>.
4. *Начкебия, М.С.* Влияние управленческой культуры на профессиональную деятельность руководителя [Электронный ресурс] / М.С. Начкебия. – 2008. – Режим доступа: [www.pacademy.edu.by/conferenc/2008/kpu/doc/razdel3/nachkebia.doc](http://www.pacademy.edu.by/conferenc/2008/kpu/doc/razdel3/nachkebia.doc).
5. Психологический словарь [Электронный ресурс]. – 2000. – Режим доступа: [http://planey.ru/dic/k/k\\_47.htm](http://planey.ru/dic/k/k_47.htm).