

Д.В.Деопик

Некоторые принципы построения формализованных языков для исследования исторических источников

1. Общая потребность в применении метода количественного анализа в исторических науках очевидна. У него есть как преимущества (возможность более полного использования источников, т.е. изучения их ранее не использовавшихся частей и более систематизированного и проверяемого анализа уже использованных частей), так и недостатки (невозможность использования обобщений, данных самими обществами, которые мы изучаем; игнорирование на первом этапе обобщений «по аналогии», полученных на основе небольшого, но признаваемого существенным материала данного источника, шире представленного в других источниках). При этом мы часто имеем дело с событиями, достаточно отдаленными во времени или пространстве, когда понятийная система была существенно отличной и не может быть полностью восстановлена. В этом состоит отличие источников исторических от экономических и демографических, где применение количественных методов развивается быстрее в силу простоты (для количественного анализа) языка самого источника.

2. Существуют два аспекта применения количественных методов в исследованиях общества: а) что считать; б) как считать.

«Как считать» – это область собственно статистики. Здесь интересно лишь то, что в наших исследованиях часты типы распределений, редкие в других науках. Примером может служить постоянный положительный эксцесс нормального распределения для временных процессов в обществе. Здесь еще мало разработаны соответствующие обще-статистические методики, а конкретные исследования дают но вый экспериментальный материал, нуждавшейся в новых статистических разработках, за счет чего всегда развивалась статистика.

«Что считать» – основная проблема при количественном анализе данных исторических источников. Необходимо привести фактический материал, чрезвычайно сложный и многосторонне взаимосвязанный, к сопоставимому количественному виду. При этом необходимо отделить, как правило при помощи громоздкой процедуры, интересующие нас явления от тех, чье влияние при данном подходе считается несущественным, и как-то оценить те, чье влияние очевидно и признается существенным. Необходимо, с одной стороны, оперировать простейшими элементами, с другой – изучить, даже в самой про стой ситуации в обществе, их сложнейшие сочетания. Решение подобных задач, задач систематики, невозможно без тех или иных процедур переработки исходной информации, и прежде всего,

переписки текстов на языке-посреднике, поскольку большой объем преобразований на пути от массы исходных фактов к интерпретации тех или иных связей и закономерностей внутри массива требует от описания каждого конкретного факта однозначности и простоты. Лишь формализованный язык-посредник может позволить систематизировать большие массивы информации и дать возможность применения количественного анализа. Помимо этого, лишь на этом языке можно проводить сопоставления данных массового материала и логически-интуитивных обобщений. Эти последние в не меньшей степени нуждаются в признанных формулировках. Будущее покажет, насколько широки области, могущие быть обслуженными одним формализованным языком, но уже накоплен (экспериментально) определенный опыт, позволяющий высказать ряд общих суждений.

3. Исключительная сложность связей между элементами, малое количество повторяющихся полностью данных, невозможность постановки активного эксперимента, наконец, отсутствие, до середины XII в, насущной практической потребности у общества в строгом анализе фактов по истории – все это привело к запаздыванию применения в истории методов, давно уже обязательных для других наук, от собственно математики до биологии. Лишь в последние десятилетия, на основе успехов социологии и развития ряда отраслей статистики (особенно теории малых выборок) формируются историометрия, археометрия и др.

Причины, которые обусловили развитие количественных методов в изучении истории общества, вызвали и многочисленность попыток уклониться от изучения материала «снизу» при помощи тех процедур, которые признаются необходимыми в таких науках, как биология. В данном случае имеется в виду отсутствие привычки к ведению громоздкой работы систематика. Это привело, во-первых, к стремлению использовать нестрогие традиционно сложившиеся обобщающие термины языка историков (археологов, этнографов, фольклористов и др.) в качестве формализованного языка. При этом игнорировалось явное несоответствие объема и характера покрываемых этими терминами фактов у разных исследователей, тем более в разное время и в разных странах. Исключение представляет история экономики ближайших к нам периодов в схожей культурной среде; здесь традиционно сложившийся язык был достаточно строгим. Во-вторых, это вызвало попытки применить на вид более строгий подход – пойти по пути построения общих моделей «сверху», исходя из накопленного наукой опыта, с последующим построением языка описания для более низких уровней. При этом предполагалось опуститься до уровня одиночного факта, но либо у исследователей не хватало времени и сил, либо получался, как в первом случае, список привычных терминов – стихийных обобщений, непригодных в качестве формализованного язычка для количественного анализа.

4. Возможна иная процедура, уже применяющаяся в той или иной степени многими исследователями, (особенно в таких областях, как археология, где степень точности описания традиционно выше, чем в собственно истории, а простейшие элементы – предметы и их элементарные связи – даны достаточно очевидно и в большом количестве, которая в данном сообщении показана на примере применения к более сложным объектам – двум основным видам исторических источников: документам и нарративным источникам.

Процедура сводится к выявлению простейших составляющих элементов для данного типа документов в его конкретном виде (т.е. в определенной культуре на определенном неслучайном временном отрезке), классификации этих элементов, рассматриваемых как признаки изучаемого явления (эволюции, по данным эпиграфического массива, аграрных отношений в Бирме XI-XV вв., эволюции, по данным летописи, межгосударственных отношений в Восточной Азии в VIII-V вв. до н.э. и пр.), и определению массовости каждого из них. На следующем этапе исследуются отношения между элементами в рамках всего периода целиком, далее – эволюция этих отношений от начала до конца периода (а также – эволюция набора элементов, т.е. появление новых и исчезновение старых элементов). В итоге мы получаем описание явления через массовые, т.е. типичные признаки, но не все, а те, которые содержатся в данном источнике или группе однотипных источников. Учет последнего обстоятельства делает для читателя очевидной основу обобщений автора и пределы приложимости его выводов.

Остановимся на двух важных чертах статистической оценки эффективности формализованных языков, точнее – их лексических элементов, являющихся признаками явления, и простейших отношений между элементами, также являющихся признаками явления. Известно, что количественное исследование начинается с выделения перечня признаков для описания, иными словами – различения типологически ближайших явлений (если перечень не задан самим текстом). Исследователю этот перечень часто кажется очень громоздким, длинным, и тем длиннее, чем большее количество признаков он берет для надежности. Естественно, что списки сочетаний значений этих признаков еще более громоздки. Описанная трудность в значительной степени преодолевается предварительным количественным анализом перечня признаков и его поведения на исследуемом пространственном или временном массиве. При этом, после оценки частоты реализации значений (в том числе нулевого) каждого признака ее для данной конкретной задачи можно не использовать:

а) значения признаков настолько редкие, что они не могут характеризовать (наличием не менее 30%) самую меньшую из естественно сложившихся групп фактов, (часть текста летописи, временной отрезок или региональная группа для эпиграфического комплекса, отдельный археологический памятник);

б) целые признаки, представленные либо одним редким значением с нулевыми другими, либо одним значением и вторым, дополняющим до 100%; в этом случае признаю также не может характеризовать ни одну конкретную естественно-исторически сложившуюся группу фактов. Такие признаки характеризуют массив в целом, но не нужны для его классификации.

Аналогично на каждом уровне исключаются из дальнейшего рассуждения все, сочетания признаков, обладающие указанными особенностями. Как показывает эксперимент, остающиеся перечни признаков (с их значениями) и их сочетания достаточно компактны.

Следующий принцип – достаточность классификации. По мере анализа образуются группы фактов, обладающие устойчивыми сочетаниями значений признаков, призванных на основании указанной выше процедуры классифицирующими. Эти группы должны сопоставляться или с набором групп, полученных при анализе на предшествующем этапе более простых сочетаний значений признаков, или с исторически данными конкретными подразделениями исследуемого массива. Задача классификации может считаться решенной в двух случаях. Во-первых, когда установлено, что сочетание значений n признаков дали группы, не сливающиеся и не дробящиеся при последовательном учете сочетаний с новыми признаками. Во-вторых, когда установлено, что сочетания значений некоторых признаков постоянно присущи некоторым исторически выделенным явлениям. В первом случае речь идет о выработке, во втором – о проверке классификации. Такова процедура сокращения (в начале и в ходе исследования) перечня признаков, дающих количественную характеристику того или иного явления или его части.

Но ситуация выработки перечня признаков исследователем типична скорее для археологии, этнографии, отчасти для исследований, связанных с языком (топонимика и др.). В собственно же исторических источниках список признаков дан текстом, и, хотя он гораздо шире при применении количественных методов, чем без него, он конечен, и исследовать его лучше целиком (с обязательной количественной оценкой).

5. Особенности составления перечня признаков для массива однородных (в различной степени) документов и одного нарративного источника целиком посвящена вторая часть настоящей статьи¹.

Текст (или интересующий исследователя аспект – экономические операции в документах) записывается в виде определенного числа высказываний (в логическом понимании этого слова). Высказывание или «простейшее событие» соответствует одному поступку человека (в эпиграфике) или акции государства (в нарративном источнике, описывающем «дела царств»). Элементы высказывания: субъект, предикат, объект, предмет действия (например: «X подарил У поле», «X захватил у У город»), а также пространственно-временная характеристика легко получаются из текста или контекста. Объем полученных простейших событий, как правило, больше, реже – равен объему соответствующей части, текста, но обладает всеми преимуществами однозначности. Но само простейшее событие – это уже отношение, довольно высокий уровень сочетания значений признаков. Сами признаки – это элементы простейшего события и как раковые они обладают определенным набором значений (различные виды субъектов и т.д.). Анализ начинается с составления перечня значений признаков, определения массовости каждого из них и получения для каждого значения (и для признака в целом) пространственно-временных характеристик. Сами эти характеристики также могут дать основу для исторических интерпретаций, разумеется, после того, как ото браны значения с достаточной массовостью.

Затем следует анализ парных сочетаний значений двух признаков, обладающих достаточной массовостью. На этом этапе выявляются не сочетающиеся значения и дается содержательная интерпретация факта несочетаемости (например, отсутствие определенных видов дарений определенным категориям лиц). Далее анализируются сочетания значений трех и более признаков, обладающие достаточной массовостью; явления, характеризующиеся такими сочетаниями, тем реже бывают массовыми, чем больше число характеризующих их признаков.

6. Последняя крупная проблема - выработка лексики формализованного языка, грамматика которого есть грамматика логических высказываний, поскольку невозможность иной грамматики при количественном подходе общепризнанна. Опыт показывает, что

¹ Подробно результаты исследований изложены в статьях Д.В.Деопик. Некоторые тенденции в социальной и политической истории Восточной Азии в VIII-V вв. до н.э. (на основе систематизации данных Чунцю). – В сб.: Китай: традиции и современность. М,1976; Д.В.Деопик. Проблемы методики исследования эпиграфического комплекса применительно к задачам социально-экономического анализа (на материале бирманской эпиграфики). – Вести. Моск. ун-та, серия востоковед.,1977, №2.

наименьшие погрешности на первом этапе исследования дает использование в качестве лексики формализованного языка самих слов текста источника. Это позволяет: а) избежать неоправданных обобщений; б) установить поле функций непереведенного или ориентировочно переведенного термина; в) проверить полученные обобщения; г) привлечь к трудоемкой работе специалистов со средним уровнем подготовки. Погрешности, возникающие при переписке «живого» текста при помощи системы логических высказываний, минимальны. Естественно, полученный переписанный текст весьма громоздок, но его дальнейшая обработка машинным способом не представляет труда, а анализ значений признаков и парных сочетаний быстро выполняется и «вручную».

7. Завершающим этапом является статистическая оценка надежности полученных соотношений на основе насколько возможно проверяемого статистически допущения о характере распределения и репрезентативности выборки. В случаях, когда исследуемый массив является случайной выборкой, сделанной исследователем, следует добавлять принятые в таких случаях оценки.

Опубликовано в сб-ке «Количественные методы в гуманитарных науках». М. 1981, с. 5-