

П.Ш. БЕРМАН

БОЛЬШИЕ ДАННЫЕ И ИСТОРИЧЕСКАЯ СОЦИАЛЬНАЯ НАУКА

БЕРМАН Питер Ш. – почетный профессор социальных наук Колумбийского университета, член Американской академии искусств и наук, член Национальной академии наук США, Нью-Йорк, США (psb17@columbia.edu).

Аннотация. Большие данные способны революционизировать социальную науку, а когда их происхождение связано с важными [историческими] контекстами, они помогают в поиске ответов на прежде не поддававшиеся решению вопросы о временных параметрах и последовательностях событий, об их границах. С учетом этого не имеет значения, относятся ли Большие данные к социальным наукам вообще или же только к науке исторической, в которой повествование основывается на нарративах (*sentences*). Поскольку такие изложения – норма, воздействие Больших данных на практику исторической социальной науки может оставаться довольно ограниченным, как ни хотелось бы достичь расширения их использования.

Ключевые слова: Big Data • Большие данные • историческая наука • историческая социология

DOI: 10.31857/S013216250008501-4

Практически в то самое время, когда исследователи сетей открывали законы и другие основанные на связях эпифеноменальные качества социальных сетей в дигитальном выхлопе онлайн-поисков, предпочтениях френдов в социальных медиа, рекомендациях покупателям, обзорах продуктов и т.п., историки потихоньку выстраивали мощные сооружения из архивных данных, почерпнутых в существующих хранилищах критически важных институтов, делая такие конструкции из данных доступными публике. Приведу всего один пример, относящийся в основном к данным по Великобритании: полный корпус текстов протоколов Олд Бейли – Центрального уголовного суда Англии и Уэльса, документы о работорговле в Атлантике, архивы британской Восточно-Индийской торговой компании. Используя простейшую исследовательскую стратегию – поиск описательных ответов, можно найти ответы практически за несколько дней, если не часов или даже минут, по тем вопросам, над которыми историк-одиночка долго корпел бы в пыльном архиве в том случае, если бы возможность обращения к Большим данным была сильно затруднена. Как эти и многие другие архивы Больших данных меняют наше понимание прошлого и тем самым нашего настоящего?

Когда такие архивы попадают во Всемирную паутину, даже при минимальных звуках фанфар, появляются материалы, подобные недавнему исследованию влияния продолжительности [светового] дня на человеческие эмоции через определение соотношений между счастливыми и унылыми твитами¹. Огромные массивы текстов, охватывающих долгие

Источник: Big Data & Society. December 2015. doi.org/10.1177/2053951715612497. Примечание журнала *Big Data and Society* о правах на перепечатку: Creative Commons CC-BY. This article is distributed under the terms of the Creative Commons Attribution 3.0 License (<http://www.creativecommons.org/licenses/by/3.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/ open-access-at-sage>).

¹Твит (от англ. to tweet – чирикать, щебетать, болтать) – короткое сообщение в социальной сети «Твиттер» длиной не более 140 символов. – Прим. ред.

периоды времени, сейчас становятся широкодоступными в цифровой форме. И неудивительно, что этот массив (опять приведу лишь отдельные примеры по США: выступления президентов «О положении страны» (State of the Union), протоколы Конгресса, протоколы решений Верховного суда)² создается в важных институтах, которые длительное время в основном сохраняли преемственность форм. Сможем ли мы теперь дать новые ответы на старые вопросы по таким старым текстам?

Аналогичным образом обстоит дело с культурными артефактами: фото, меню ресторанов Нью-Йорка, планы кресел в зале Нью-Йоркской филармонии, ранее также в основном недоступные, становятся все более открытыми в онлайн-овых хранилищах [Accominetti, Khan, 2015]. Большая часть этих данных несет на себе печать времени; многие из них могут к тому же быть четко привязаны к месту. А это называется геореференцией. Сделают ли эти новые депозитарии относящихся к культуре свидетельств возможным изменение понимания критически важной динамики – формирования элит, интеграции этнических групп, распространения заболеваний, даже долгосрочных изменений климата?

А мы ведь находимся в самом начале архивной революции, исторические архивы особой важности быстро переводятся в цифру. И что удивительнее: имеются еще и новые архивы пока не выявленной важности. Они огромны по массивам документов. К примеру, два миллиарда e-мейлов, ежегодно поступавших в Госдепартамент США (и исходивших из него) при Клинтоне, несопоставимы с количеством оригиналов документов, исходивших из администрации А. Линкольна за все время его президентства³. Значимо ли что-либо из этого потока [электронной почты] для трудов историков и специалистов по исторической социальной науке? – Не удивляйтесь, что ответ на этот вопрос таков: и да, и нет.

Голдер и Мэси [Golder, Masy, 2011], писали: «Сеть все видит и ничего не забывает». Как некий фантастический образ эта фраза не сильно отличается от идеала летописи, которая бы фиксировала все «так, как оно происходило». Придумал эту формулу в 1968 г. Артур Данто [Danto, 1968], поставивший вопрос: изменит ли по существу что-нибудь подобная летопись в том, как и о чем пишут историки? Данто признавал, с учетом всего того, чем историки занимаются в архивах, что наличие такой летописи позволит им моментально проверять свои факты. Однако это – в лучшем случае – тривиальный элемент их практики. Что делают историки – и только они? Они творят нарративы, например «в канун Рождества 1642 г. родился отец современной физики». Летопись, фиксирующая все так, как что-то произошло, к такому роду деятельности полностью непригодна, ибо лишь после того, как родилась современная физика, Ньютон мог бы стать ее отцом. Ткань летописи, видевшей все и ничего не забывающей, абсолютно точно констатировала бы: в канун Рождества 1642 г. родился Исаак Ньютон, но никогда бы не стала констатировать в тот момент, что несколько столетий спустя Ньютон будет назван отцом современной физики. То есть наши новые структуры данных не имеют большого значения для центрального элемента ремесла историков: писать повествования-нарративы.

Но для истории как науки социальной, не занимающейся писанием подобных нарративов, такие структуры данных могли бы стимулировать – в широком диапазоне важных, значимых проблем – радикально иное представление о прошлом, а следовательно, и некую реорганизацию нашего понимания настоящего, ибо новые виды данных вызывают к жизни новое видение истории и обеспечивают его аргументами, – такой истории, которая сосредоточена на динамике, распознавании паттернов и идентификации механизмов, посредством которых действия акторов, накапливаясь, генерируют результаты макроуровня. Речь идет о наших самых сложных интеллектуальных проблемах – например, о чем-то похожем на возникновение «общества модерна» (modern society). Что могут здесь внести Большие данные в наше понимание?

² Документы Конгресса США см.: <http://thomas.loc.gov/home/LegislativeData.php?n=Record>; «О положении страны» см.: <http://www.presidency.ucsb.edu/sou.php>; протоколы Верховного суда см.: <http://www.loc.gov>.

³ Я благодарен Матту Коннелли за эту информацию.

Хосе Атриа, работающий (2015) над PhD-диссертацией в Колумбийском университете, пишет об использовании архива *Олд Бейли* в поисках ответа на вопрос именно такого рода. Отправной точкой проекта Атриа стало следующее его наблюдение: историки и обществоведы (social scientists) сходятся в том, что целый комплекс процессов, начала которых многими видятся в XVI в., привел к возникновению современного (modern) общества. В этом нет сомнений. Атриа писал: «Они включают такие перемены, как возникновение разновидностей образа жизни, сделавших насилие и тело отталкивающими и постыдными [Elias, 2000], как рост мощи централизованного государства и возникновение национальных общностей за пределами непосредственно родственных связей [Bearman, 1993; Anderson, 2006; Tilly, 1975], развитие рыночного общества [Polanyi, 2001], распад традиционных социальных отношений, основанных на статусе, в пользу социальной структуры, базирующейся на общественных классах [Thompson, 1963], и возникновение дисциплинирующих методов контроля и власти [Foucault, 1977]» [Atria, 2015].

К этому есть, конечно, что добавить. Но и сказанного достаточно для нашей задачи. Главный вопрос ее, не что могут быть и иные [процессы], а в том, что в отношении структуры и наличия каузальных связей⁴ между ними мало согласия. Ответ на такой вопрос (а его можно разбить на целый ряд важных подвопросов, например: как все эти полуавтономные процессы, внося свой вклад в возникновение общества модерна, взаимодействуют один с другим? Какие перемены были первыми? Какой процесс следовал [за предыдущим] по времени своего начала, по скорости перемен?) дает нам как раз то, с чего в игру могут вступать Большие данные – в нашем случае архивы *Олд Бейли*.

Важно, что эти архивы сложились непосредственно в институции, которая была абсолютным центром всех тех критически важных процессов, которые генерировали возникновение *модернити*. Но и другие институции играли столь же важные роли. Часто отмечаемый характер исторических данных (он верен и для больших исторических данных, обещающих изменить наше знание о каузальной динамике в контексте истории) таков: она (динамика) возникает в контекстах институтов, основательно вовлеченных в эти процессы. Сохраненные вещи – чаще, чем нет, – это вещи, которые тогда люди считали нужным хранить. Архив, охватывающий период с 1674 по 1913 г. (то есть в конце концов период, в границах которого складывается *модернити*), содержит точные, с зафиксированным временем, данные по миллионам уникальных личностей, тысячам мест, сотням тысяч интеракций – и криминальных, и повседневных. В записи (документы) внесены точные данные, отражающие структурные отношения, позволяющие нам сосредоточиваться на меняющихся моделях (паттернах) социальных отношений, на социальной географии насилия, отступлении фактора родства, на распаде разного рода местничества, появлении новых стратегий и дискурсов о власти и управлении (governance). На основе таких данных мы можем идентифицировать скорость институциональных перемен, каузальную динамику – основу сложного узла взаимозависимых и связанных друг с другом процессов, конституирующих *модернити*. Или – как минимум – можно ожидать, что подобный анализ – по-другому он неосуществим – в пределах досягаемости.

А возьмите схожую проблему: понимание последовательности исторических событий – точнее: как нам выявлять периоды? Ставшие с недавних пор доступными хранилища текстов охватывают длительные периоды времени, обеспечивая надежные предпосылки идентификации поворотных пунктов истории. Как и в дискуссиях прежних времен, американские историки согласны, что в такое-то время американский социальный и политический дискурс относительно хода истории за более чем два века – 1790 г. по настоящее время – изменился. Они согласны и в отношении природы произошедших перемен,

⁴Я говорю о каузальности в широком веберовском смысле термина. К примеру, «Протестантская этика и дух капитализма» имеет в виду одну сторону каузальной цепочки, соединяющей дух капитализма с аскетическим протестантизмом, а не более узкую версию этого термина, распространенную в науке сейчас.

согласны в том, что современный дискурс можно с уверенностью отличать от его прежних форм. Но вопросы: когда эта перемена произошла и как нам оптимально ее идентифицировать? – вызывают серьезные дебаты. Канонические объяснения в центр ставят 1865, 1877, 1896 и 1913 годы. И почти каждая такая дата имеет своего сторонника! Такие споры по необходимости порождают стратегии ученых в конкретном понимании истории. То есть порождают стратегии идентификации отрезков фактически непрерывного потока уходящего прошлого, достаточных для осмысленного выделения отдельных сгустков времени, таких как «позолоченный век», «прогрессивный период», «век фронта (frontier)» и т.д. Периодизация отрезков времени, выбираемая нами, зависит, конечно, от исходной точки зрения. История Америки, страны иммигрантов, порождает периодизацию, отличную от истории Америки – рыхлой конфедерации отдельных штатов. С учетом сказанного вызов таков: создать наделенную смыслом «нарезку» периодов между модерном и до-модерном, наполнив их социально-политическим дискурсом, при которой содержанием такого дискурса выступали бы именно риторические инструменты периодизации! Чтобы совершить это, требуется идентифицировать форму, которая преемственна и значимость которой не прерывалась. Здесь-то и выходят на сцену Большие данные.

Опираясь на архив президентских посланий о положении страны, ряд авторов [Rule et al., 2015] показали, что послания президентов обеспечивают, несмотря на разные формы выступлений, одну из подобных форм, тем самым дав нам Архимедову точку опоры и позволив увидеть содержание, структуру американского социально-политического дискурса, развертывающегося во времени.

Используя новые стратегии анализа исторических текстов, где термины, концепции, способы использования языка и смыслы слов меняются во времени, названные выше авторы показывают: современный (модерный) социальный и политический дискурс о природе управления и власти появляется как особый объект *после* 1917 г., хотя элементы подобного дискурса можно идентифицировать и ранее. Тем самым мы получили новые возможности проникновения в понимание истории США. В этом же русле трансформируется и наше понимание того, как выглядит дискурс модерна и пред-модерна, и, что не менее важно, элементов преемственности, их связывающих. Попутно они демонстрируют новую стратегию идентификации смысловых категорий в корпусе текстов, охватывающих продолжительные периоды времени. И еще: такой подход позволяет раскрыть текучесть категорий дискурса во времени и анализировать их преемственность, идентифицируя разговорный поток как объект [научного] интереса. А то, что он – этот подход – дает возможность получать визуализации, которые позволяют идентифицировать перемены лексиксы за длительные промежутки времени (рис.), – это еще один полезный результат.

Целый класс новых проблем, которые пока не поддаются точной формулировке, может стать доступным благодаря Большим данным. Используя новую структуру данных, точно привязывающую к месту (geo-reference) держателей сезонных абонементов Нью-Йоркской филармонии по параметрам места их кресла в зале (!) и места жительства, Акоминетти и Хан [Accominetti, Khan, 2015], например, показали, как элита Нью-Йорк Сити использовала меценатство в отношениях с институтами культуры для получения признания и как эта возможность взаимопризнания вела к последовавшей реорганизации моделей расселения в центре города в «позолоченный век». Одна из находок здесь такова: элитарный класс стал смотреть на себя через призму взаимного соблюдения привязки к ряду районов, микрорайонов, концертов и т.д., в то же время давая неэлитам возможность циркулировать на границах элитного субмира.

Еще пример. Хоффман и др. [Hoffman et al., 2015], используя данные Архива работорговли в Атлантике, дополнили их детальными данными о занятости, взятыми из ставших недавно доступными геоферентными указателями улиц Ливерпуля. Они смогли показать, почему ливерпульская сеть работоторговцев устояла под атаками пиратов и «победила» Лондон с Бристолем в борьбе за контроль над торговлей в треугольнике Англия – Северная Америка – Западная Африка после 1750 г. Мюллер и его коллеги [Muller et al., 2015]

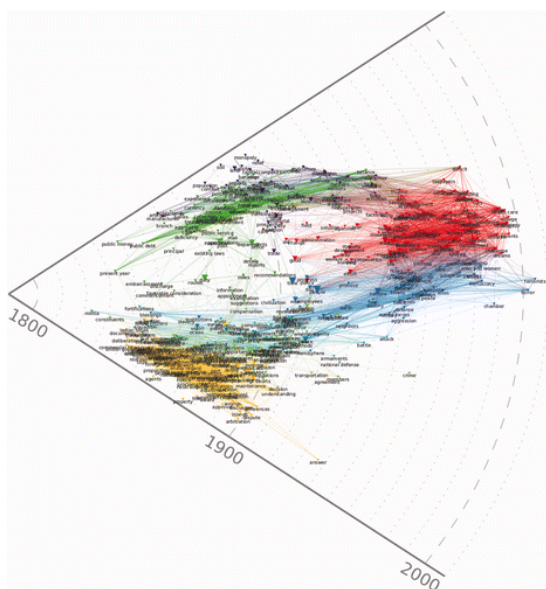


Рис. Радарная схема президентских посланий о положении страны: слова привязаны к определенному времени при 50% случаев повторения

аналогичным способом переводят в цифру данные о назначении сроков тюремных заключений в г. Лоуренс, штат Массачусетс во время знаменитой стачки «Хлеб и Розы». Все это – в сочетании с другими административными данными (переписи, смертность, больницы) – позволяет смоделировать диффузию сопротивления в одном из канонических событий истории трудящихся Америки. Теоретически все эти исследования были возможны и до появления Больших данных. Но ни одно из них не было бы завершено в течение жизни отдельного ученого. Перспективность таких исследований и в том, что они вносят вклад в изучение конкретной проблемы истории, которой эти ученые занимаются, и в идентификацию механизмов, которые можно переносить и на другие контексты. Использование Больших данных для идентификации переносимых механизмов, думаю, – это главное. Иначе мы останемся просто с описанием, что нужно, но не очень полезно социологически.

Приведенные выше примеры лишь поверхностно коснулись тех возможностей, которые открывает революция Больших данных перед исследователями истории. Целый космос работ, связанных со структурами административных данных (например, переписи), институциональными документами (например, сведения по штатам о де-институционализации умственно больных в период 1975–1980 гг.); данных, связывающих стандартные инструменты социальных наук с крупными историческими архивами (например, GSS – Всеобщая социальная перепись – с данными о смертности в США и т.д.), здесь не были затронуты. Они по своим характеристикам аналогичны описанным выше проектам: ориентация на получение ответов на каузально важные вопросы посредством идентификации переносимых механизмов. Вместо написания нарративных изложений фокус сделан на идентификации ритма и структур перемен, проистекающих из ключевых институциональных точек, которые, как известно, выступают основой крупных и интересных исторических процессов. Эти структуры Больших данных могут произвести подлинную революцию в исторической социальной науке. Однако, затрагивая лишь эпифеноменальные материалы, относящиеся к не-центральной институциональным точкам, Большие данные не кажутся очень многообещающими. В конце концов, что на входе, то и на выходе.

СПИСОК ЛИТЕРАТУРЫ [REFERENCES]

- Accominetti F., Khan S. (2015) *The NY Philharmonic Project*, Mellon Foundation. URL: <http://incite.columbia.edu/subscribers-to-the-ny-philharm> (accessed 10.01.2015).
- Anderson B. (2006) *Imagined Communities*. London: Verso.
- Atria J. (2015) *Dissertation Proposal*, Columbia Sociology. Unpublished Manuscript.
- Bearman P. (1993) *Relations into Rhetorics: Elite Social Structure in Norfolk England, 1540–1640*. New Brunswick, NJ: Rutgers University Press.
- Danto A.C. (1985) *Narration and Knowledge: Including the Integral Text of Analytical Philosophy of History*. New York: Columbia University Press.
- Elias N. (2000) *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Revised edition. Oxford: Blackwell Publishing.
- Eltis D. A *Brief Overview of the Trans-Atlantic Slave Trade Voyages: The Trans-Atlantic Slave Trade Database*. URL: <http://www.slavevoyages.org> (accessed 10.01.2015).
- Foucault M. (1977) *Discipline & Punish*. New York: Random House Digital, Inc.
- Golder S.A., Macy M.W. (2011) Diurnal and Seasonal Mood Vary with Work, Sleep, and Day-length across Diverse Cultures. *Science*. No. 333(6051): 1878–1881.
- Hitchcock T., Shoemaker R., Emsley C. et al. *The Old Bailey Proceedings Online, 1674–1913*. Version 7.0. URL: www.oldbaileyonline.org (accessed 24.03.2012).
- Hoffman M., Makovi K., Bearman P. (2015) *The Structure of the Atlantic Slave Trade Ownership Network*. (Unpublished manuscript).
- Klingenstein S., Hitchcock T., DeDeo S. (2014) The Civilizing Process in London's Old Bailey. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 26. No. 111: 9419–9424.
- Muller S.C.S., Makovi K. et al. (2015) *The Lawrence 'Bread and Roses' Strike, RWJ-HSS Proposal*. (Unpublished manuscript).
- Polanyi K. (2001) *The Great Transformation: The Political and Economic Origins of Our Time*. 2nd ed. Boston, MA: Beacon Press.
- Rule A.R., Cointet J.P., Bearman P.S. (forthcoming) Lexical Shifts, Substantive Changes, and Continuity in State of the Union Discourse, 1790–2014. *Proceedings of the National Academy of Sciences of the United States of America*, PNAS early ed. URL: www.pnas.org/cgi/doi/10.1073/pnas.1512221112 (accessed 30.06.2015).
- Thompson E.P. (1963) *The Making of the English Working Class*. London: Victor Gollancz.
- Tilly Ch. (1975) *Coercion, Capital, and European States, AD 990–1992*. Cambridge, MA: Blackwell.

Перевод с англ. Н.В. РОМАНОВСКОГО

РОМАНОВСКИЙ Николай Валентинович, д. ист. н., гл. науч. сотр. Института социологии ФНИСЦ РАН; проф. РГГУ, Москва, Россия (socis@isras.ru)

BIG DATA AND HISTORICAL SOCIAL SCIENCE

BEARMAN P.S.

Columbia University, USA

Peter S. BEARMAN, Cole Professor of Social Science in the Department of Sociology, Columbia University, New York, USA (psb17@columbia.edu).

Abstract. 'Big Data' can revolutionize historical social science if it arises from substantively important contexts and is oriented towards answering substantively important questions. Such data may be especially important for answering previously largely intractable questions about the timing and sequencing of events, and of event boundaries. That said, 'Big Data' makes no difference for social scientists and historians whose accounts rest on narrative sentences. Since such accounts are the norm, the effects of Big Data on the practice of historical social science may be more limited than one might wish.

Keywords: Big Data, historical social science.

Transl. by N.V. ROMANOVSKY

Nikolai V. ROMANOVSKY, Dr. Sci. (Hist.), Chief Researcher, Institute of Sociology of FCTAS RAS; Prof., Russian State University for the Humanities, Moscow, Russia (socis@isras.ru)