

РАССУЖДЕНИЕ О МЕТОДАХ

А. А. ТОКАРЕВ

Big Data в исследовании соцсетей: опыт неудачного машинного анализа украинского Facebook*

Алексей Александрович Токарев, канд. полит. наук, ст. науч. сотр.
Центра глобальных проблем Института международных исследований
МГИМО МИД России. E-mail: a.tokarev@inno.mgimo.ru

Аннотация. В статье представлено подробное описание первого этапа работы межвузовской команды из МГИМО, МГУ, Института экономики РАН по исследованию украинского сегмента социальной сети Facebook в части изучения дискурса украинских лидеров общественного мнения. Исходя из убеждения в том, что именно Facebook на Украине является основным каналом коммуникации политиков с населением, и собрав базу данных из 176 аккаунтов представителей элиты (а затем – и постов, написанных в них за 10 месяцев), исследователи запустили машинную обработку данных. Исследовательский вопрос был сформулирован следующим образом: «Какие стратегии государства в отношении конфликта на Донбассе вербализируются украинскими элитами?» В процессе работы над полученными автоматизированным путем данными автор столкнулся с явными ограничениями машины, которые были порождены, с одной стороны, неудачной операционализацией понятий в системе предмета исследования, с другой – несовершенством самого механизма «Семантический архив», ожидания от которого оказались завышенными, с третьей – непониманием автора специфики работы с «большими данными». Однако именно неудачный опыт пилотажного исследования позволил поставить принципиально важные вопросы для дальнейшей работы – прежде всего, о критериях базы данных и об исследовательских вопросах, которые автор адресует машине. Этот неудачный опыт оказался крайне важным для второй и третьей волн исследования, завершено спустя полтора года после его начала. Поэтому рефлексия в отношении такого опыта, по мнению автора, должна быть предана гласности.

Ключевые слова: большие данные, Facebook, big data, Украина, конфликт в Донбассе.

В этой статье рассказано об опыте неудачи в пилотажном исследовании, во многом благодаря которому две следующие волны оказались успешными. Сначала представлен краткий об-

*Автор благодарит за помощь в работе Адлана Маргоева и Дмитрия Нилова (МГИМО), Дарью Щеглову (МГУ) и Максима Бороденко (Институт экономики РАН).

зор литературы по теме, после этого указано на сложности проведения операционализации для постановки технического задания программистам, затем дается ответ на вопрос: почему для исследования мнения украинских элит был выбран именно Facebook, рассказано о формировании списка аккаунтов и работе с поисковым запросом. Наконец, обсуждается причина, по которой машина не смогла справиться с поставленной задачей, и характеризуются вопросы, которые в результате возникли при переходе к следующим этапам исследования.

«Большие данные» используются в количественных исследованиях социальной реальности на Западе более 10 лет [13; 14]. При этом исследований в отношении украинского конфликта на примере соцсетей мы не встречали. Впрочем, во многих работах эта тема заявлена, но либо вообще никак не раскрывается (несмотря, например, на многообещающее заглавие [11]), либо раскрывается не полностью. Так, в статьях А. Ронжина [16], П. Н. Олещук [8] и Г. Ю. Никипорец-Такигава [7] постулируется значимость социальных сетей в современном мире, показана их возрастающая роль как средства коммуникации и информации, а также политической мобилизации общества, но не исследована специфика украинского сегмента ни Facebook, ни Twitter.

Тема украинского конфликта представляет интерес как для состоявшихся ученых из разных областей науки, от филологии до истории и политологии, так и для тех, кто только вступает на академическую стезю. Нам удалось обнаружить курсовые и дипломные работы, посвященные реакции СМИ и интернет-пользователей на конфликт [10].

Одно из исследований российских ученых было посвящено изучению протестного поведения в соцсетях во время Евромайдана в 2013 г. [1]. Помимо изучения собственно протестного поведения, российские ученые объяснимо фокусируют свое внимание на анализе российских взглядов на украинский конфликт и присоединение Крыма [6] и исследовании изменения отношения к России, которое можно проследить по постам [5], репостам [9] и комментариям [4] в социальных сетях и блогосфере.

Общий обзор той области исследований, что получила название «социального компьютеринга», можно найти в работе российского социолога А. А. Давыдова [3]. В России на данном направлении заметны работы коллег из Института социологии РАН, МПГУ и Санкт-Петербургского университета информационных технологий, механики и оптики [2]. Одна из последних масштабных работ, объясняющих методы «социальных вычислений», издана Л. З. Мановичем в 2015 г. [15]. Последний по времени всплеск интереса широкой общественности к «большим данным» относится к концу 2016 г. и связан с освещением использования Big Data командой Дональда Трампа в рамках президентской кампании [12].

В ожидании чуда: почему для работы с машиной необходима адекватная операционализация?

Работа с автоматизированной системой анализа и поиска информации виделась маленьким волшебством. Одни объемы в сотни тысяч и миллионы единиц анализа, с которыми ассоциируются слова Big Data, чарующе действуют на исследователя-гуманитария, не избалованного количественными методами, особенно в современной политической науке в России.

Казалось, процессы собирания того, что украинские элиты пишут по поводу населения Донбасса, и анализа этих сотен дискурсов будут похожи на поиск смартфона на Яндекс.маркете. Нельзя сказать, что мы пришли к разработчикам с праздной просьбой — «мне нужен телефон — помогите выбрать» («нам бы поисследовать дискурсы украинских элит в отношении конфликта»). Определенную операционализацию в системе предмета исследования мы провели — поставили исследовательский вопрос, сформулировали гипотезы, наметили основные методы. Но рассчитывать на их «фильтрационную» помощь («А сколько у вас денег на телефон? Диагональ хотите большую? Камеру хорошую? Оплачивать телефоном покупки хотите?») не приходилось. В отличие от продавцов телефонов, программисты ничего не знают про украинское национальное строительство, донецкий сепаратизм, конфликтные стратегии акторов, постимперский транзит и прочее. Поэтому каждый раз при общении их просьбы сводились к вежливому фразам: «пожалуйста, опишите точнее, что вы хотите исследовать?».

Процесс общения исследователя с машиной облегчается многократно, когда исследователь четко формулирует, чего он хочет от нее. Но на этапе знакомства возникают сложности. Кого считать украинским блогером? Живущий в Гааге, но влияющий на общественное мнение,

вероятно, не меньше топовых аналитических программ Анатолий Шарий — украинский блогер? Что такое «украинские элиты»? Доктор Комаровский (в 3,5 раза больше подписчиков, чем у премьер-министра Украины), певицы Ани Лорак (в 3 раза) и Вера Брежнева (в 2,2 раза) — часть элит? Они формируют смыслы? Так ли уж стоит восхвалять украинский сегмент Facebook в качестве объекта для репрезентативного исследования, когда туда не пишут бизнесмен Игорь Коломойский и директор президентского НИСИ Владимир Горбулин? Каковы границы Украины, интересующей нас в исследовании?

И что делать с цитатами? Человек понимает разницу между дегуманизирующей фразой — «Эти люди должны сдохнуть!» и миротворческим вопросом — «Как можно говорить о согражданах: “Эти люди должны сдохнуть!”?». Но автоматизированная обработка, наткнувшись на обусловленные нами маркеры, заберет оба поста в одну корзину. Как объяснить силу кавычек машине?

При всем обаянии Big Data воображать, что мы можем исследовать всё, просто глупо. Во-первых, огромные объемы единиц анализа работают против нашей скрупулезности. Невозможно бравурно считать, что за рамками 176 блогеров не нашелся кто-то, кого мы не учли. Во-вторых, выбрать телефон много проще, чем исследовать огромный социальный феномен при помощи Big Data. Во втором случае тема размыта. Сколько синонимов у смартфона Xiaomi Redmi Note 3 Pro? Ноль. Сколько можно придумать слов-маркеров, связанных топонимически, исторически и социокультурно с топонимом «Донбасс», заставляющих машину положить пост в нужную нам корзину, даже если «Донбасс» там не употребляется? Начинали с трех, спустя полтора года загружали выборку из 110.

Нет волшебной кнопки, которая, будучи описанной многостраничным программным кодом в соответствии с нашими пожеланиями и проведенной операционализацией, в один миг (клик!) выдаст ответ на поисковый запрос настолько точный, что именно то, что мы искали, будет в базе данных, а то, что нам принципиально не нужно, туда не попадет.

Почему украинский Facebook?

До сих пор при обсуждении докладов в рамках нашего проекта приходится слышать от старших коллег «этот ваш Интернет...» и «какой-то там Facebook...». С одной стороны, очевидным и не требующим доказательств кажется тезис о многократно возросшей выросшей роли Интернета (и социальных сетей, в частности) в общественной жизни, в особенности после событий «арабской весны». Оценка руководителя Роскомнадзора А. А. Жарова отражает реальность: «Как показывает практика организации массовых беспорядков в России, в том числе прошлогодние события в Западном Бирюлеве, на Манежной площади, или зарубежные примеры — в Турции, в Великобритании, на Ближнем Востоке, в Греции, на Украине, — социальные сети играют организующую роль»¹. Другое дело, что чиновники нередко воспринимают Facebook как источник всех бед, хотя он — не более чем инструмент. Во-вторых, читателю, находящемуся вне украинского контекста, необходимо обстоятельно ответить на поставленный выше вопрос.

В России социальные сети политиков являются вторичным инструментом коммуникации по сравнению с их интервью и сообщениями «источников» на телевидении и в прессе. Исходя из этого, топ-политики крайне редко ведут социальные сети самостоятельно, отдавая это на откуп пресс-службам. Что, впрочем, не исключает двойной ответственности за аккаунт, как, например, у глав Чечни и Татарстана, которые сами часто выкладывают в Instagram сторизы² и селфи, но чьи посты, как правило, готовят сотрудники.

В отличие от России, где всплеск политического интереса к соцсетям был ситуативно связан с особенностями лидерства президента Д. А. Медведева, Украина — государство, где власть и лидеры общественного мнения в буквальном смысле живут в Facebook, который является одним из важнейших инструментов репрезентации их для общества в целом и избирателя в частности. С одной стороны, эта страна, как и Россия, находится в общемировом тренде медленной смерти традиционных медиа в связи с большей оперативностью соцсетей в формировании новостного потока. С другой — Facebook является на Украине основным каналом коммуникации внутри

¹ Мухаматулин Т. 10 лет назад появилась социальная сеть Facebook. Отдел науки приводит мнения ученых о Facebook и социальных сетях // Gazeta.Ru. 2014ю 4 февраля. — URL: https://m.gazeta.ru/science/2014/02/05_a_5880181.shtml

² Короткие самоудалющиеся видео, «живущие» в сети ровно сутки.

гражданского общества, особенно в его «активистском» сегменте. Кроме того, это одно из основных средств донесения позиции от власти к гражданам. Поэтому многие топовые политики ведут аккаунт самостоятельно, что делает Facebook еще более важным каналом коммуникации и формирования новостного потока (о многих политических событиях украинцы узнают именно из социальной сети, а не из традиционных медиа). Речь идет именно о Facebook, а не о Twitter, Instagram и других социальных сетях. На Украине было несколько попыток создать национальные социальные сети. Вероятно, самой известной из них стал канадский стартап Ukrainians, который закрылся через несколько месяцев после создания в мае 2017 г.³ Украинское общество в отношении важнейших государственных вопросов в высокой степени опосредует украинскую политику, являясь ее важнейшим участником. Facebook можно назвать в этом смысле зеркалом: если события или процесса не видно в этой соцсети, значит, к нему никто не готовится, а в нашем случае — ни один из множества планов реинтеграции Донбасса оказался не замечен блогосферой в долгосрочной перспективе.

Facebook был главным каналом коммуникации на Евромайдане⁴. Собственно, с призыва украинского журналиста первый, мирный Евромайдан в ноябре 2013 г. и начался. Как пишут сами украинцы, «в тот вечер пост собрал более тысячи лайков, около 900 репост, 1 200 комментариев и несколько тысяч людей на Майдане. Сейчас такое количество реакций на пост Найем считается обыденностью. Но до этого украинские блогеры и журналисты в среднем получали не более 100 “лайков” и десяток комментариев»⁵. По итогам тех событий вышло несколько книг, посвященных именно этой социальной сети. Часть из них представляли собой распечатки постов и дискуссий из Facebook. Даже АП Украины, рассылая журналистам просьбы «подсветить» ту или иную тему, особо акцентируют внимание на Facebook.

Формирование списка аккаунтов и работа с поисковым запросом

При постановке задач и отработке методологии мы помнили, что а) Facebook не является единственным каналом коммуникации с населением; б) мы не охватываем всех интеллектуалов, формирующих политическую повестку, и всех лиц, принимающих решения; в) даже в рамках обозначенного нами барьера в 10 000+ подписчиков в итоговый список вошли не все блогеры; г) далеко не все блогеры вели аккаунты на протяжении всего исследуемого периода, что обуславливает неравномерное распределение общего дискурса по годам (чем позже, тем больше объем); д) количество подписчиков — непостоянная величина, кроме того, не все из них — уникальные пользователи, поскольку часть является ботами.

На первом этапе при помощи мозгового штурма мы набросали список тех украинских лидеров общественного мнения, которых смогли вспомнить, получив 116 позиций. Потом мы разослали этот список коллегам с просьбой верифицировать, что увеличило его до 135. В дальнейшем мы использовали различные украинские неакадемические публикации, посвященные составлению всевозможных «топов» украинского сегмента соцсети⁶, — на первом этапе в исследовании использовалось 176 аккаунтов. Здесь мы выставили барьер в 10 000 подписчиков: аккаунты, имевшие меньшее число, в список не включались. Почему именно 10 000+? Это

³ 7 украинских социальных сетей // AIN.UA. 2017. 19 мая. — URL: <https://ain.ua/2017/05/19/7-ukrainskix-socialnyx-setej>

⁴ В официальном украинском дискурсе эти события и последовавший за ними государственный переворот называются «Революция достоинства».

⁵ Анатомія українського Facebook. Боти політиків, рейтинги, перемога на виборах // Еспрессо. 2017. 14 декабря. — URL: https://espreso.tv/article/2017/12/13/retyng_ukrayinskogo_facebook

⁶ Исследованиями эти журналистские материалы мы называем с очень большой натяжкой. В отсутствие серьезных академических проектов, посвященных украинской блогосфере, мы пользуемся тем, что есть в публичном доступе: Рейтинг самых популярных аккаунтов политиков и блогеров Украины в Facebook в августе 2015 года // Politolog.net. — URL: <http://politolog.net/ukrnews-php/rejting-samux-populyarnyx-akkauntov-politikov-blogerov-ukrainy-v-facebook-za-avgust-2015-g/>; Рейтинг самых популярных аккаунтов политиков и блогеров Украины в Facebook в мае 2015 года // Биржевой лидер. — URL: http://antikor.com.ua/articles/45909-nazvany_populjarnye_akkauntyni_politikov_i_blogerov_ukrainy_maja_2015_g_v_facebook; ТОП-50 украинских блогеров в Facebook и Twitter. Рейтинг НВ // Новое время. 2014. 1 декабря. — URL: <http://nv.ua/publications/virtualnye-geroi-politiki-i-zhurnalisty-tesnyat-pop-znamenitostey-v-facebook-i-twitter-infografika-nv-22945.html>; Померялись Фейсбуком: кто из политиков лучший блогер в 2015-м // Ліга.Бизнес. 2015. 9 июля. — URL: <http://biz.liga.net/all/it/stati/3056600-pomeryalis-facebookom-kto-iz-politikov-luchshiy-bloger-v-2015-om.htm>; 100 блогеров 2015 года. Рейтинг ICTV-факты. — URL: <http://bloggers.fakty.ictv.ua/>; ТОП рейтинг блогеров // UAinfo. — URL: <http://uainfo.org/rating>

было аксиоматическое решение — с тем чтобы получить хоть какой-то порог, аккаунты выше которого считались бы влиятельными.

После того как был определен список блогеров и сформирована база данных, встали конкретные вопросы: (1) как из общего дискурса вычлениить «куски», посвященные Донбассу, (2) как их исследовать, чтобы понять отношение украинских элит к территории и населению Донбасса.

Основным инструментом, с помощью которого проводилось изучение базы данных, был поисковый запрос. Следовательно, задача сводилась к тому, чтобы сформировать такой поисковый запрос, который: а) будет охватывать максимум постов, посвященных Донбассу, б) будет исключать те, которые не имеют к нему отношения. Задача облегчалась тем, что нас интересовало отношение к региону в целом, а не только на этапе войны.

Начиная с самых простых слов, которые первыми приходят на ум, мы все время усложняли поисковый запрос. Первоначально он состоял всего из трех слов — Донбасс, ДНР, ЛНР (с расшифровкой аббревиатур, в том числе). При каждой последующей итерации добавлялись синонимы. После добавления уничижительных слов («Луганда» и «Донбабве»), при помощи которых в 2014–2015 гг. украинская блогосфера активно дегуманизировала население Донбасса, количество постов выросло незначительно. Чуть больше массив вырос после включения аббревиатуры официального дискурса Украины в отношении непризнанных республик — ОРДЛО (отдельные районы Донецкой и Луганской областей). Максимально на первом этапе БД обогатили синоним «Минск...»⁷ и перевод синонимов на украинский. Совокупность из 14 синонимов выдала 5 200 записей из общего числа в 88 536 постов. Уровень упоминания объектов поиска в совокупности из 5 200 постов (везде с учетом морфологии), представлен в *Табл. 1*.

Таблица 1

Работа с поисковым объектом

Объект поиска	% от числа постов, посвящённых Донбассу (5 200)
«ОРДЛО»	4,1
Перевод всех синонимов на украинский	10,8
«Минск...»	26,5
«Донбас...», «ДНР», «ЛНР» и расшифровки аббревиатур	76,7

В дальнейшем с увеличением числа синонимов росло и число найденных нами в общей совокупности постов, посвященных региону (см. *Табл. 2*). В итоге удалось добиться роста с 4,5 до 5,87 %, с добавлением к первоначальным пяти синонимам еще девяти.

Таблица 2

Работа с синонимами для поискового запроса

Количество синонимов	Поисковый запрос	% от совокупности всех постов
5	«Донбас*» OR «ДНР» OR «ЛНР» OR «Донецк* народн* республик*» OR «Луганск* народн* республик*»	4,5
7	«Донбас*» OR «ДНР» OR «ЛНР» OR «Донецк* народн* республик*» OR «Луганск* народн* республик*» OR «Луганд*» OR «Донбабве»	4,54

⁷ «Минские соглашения» не прописываются в запросе отдельно, поскольку являются составной частью морфологии запроса «Минск...». Сюда, конечно, попадали и те посты, которые упоминали «Минск...» и без привязки к Минским соглашениям — например, отчеты о поездках в белорусскую столицу. Этим неудобством было решено пренебречь, ввиду незначительности в массиве.

8	«Донбас*» OR «ДНР» OR «ЛНР» OR «Донецк* народн* республик*» OR «Луганск* народн* республик*» OR «Луганд*» OR «Донбабве» OR «Минск»	5,17
9	«Донбас*» OR «ДНР» OR «ЛНР» OR «Донецк* народн* республик*» OR «Луганск* народн* республик*» OR «Луганд*» OR «Донбабве» OR «Минск» OR «ОРДЛО»	5,24
14	«Донбас*» OR «ДНР» OR «ЛНР» OR «Донецк* народн* республик*» OR «Луганск* народн* республик*» OR «Луганд*» OR «Донбабве» OR «Минск» OR «ОРДЛО» OR «Мінськ*» OR «Мінськ* домовленност*» OR «Донецьк* област*» OR «Луганськ* област*» OR «окремі райони Донецької і Луганської областей»	5,87

Итоговый запрос представляет из себя совокупность 14 синонимов, каждый из которых ищется машиной с учетом морфологии, что обозначается знаком (*) (то есть поиск учитывает только основу слова, не зависящую от окончания). Кроме того, учитывая билингвальность украинского Facebook, мы перевели все синонимы на украинский, что также учитывалось машиной. В Табл. 2 те слова, которые имеют на обоих языках одинаковые основы, но возможные разные окончания, обозначены Донбас* (рус., укр.). Слова с разными основами писались на обоих языках: Минск* соглашен*, Мінськ* домовленност*.

Провал в определении стратегий

В октябре 2016 г. в самом начале работы над проектом гипотеза формулировалась так: «При использовании метода анализа “больших данных” удастся реконструировать семантическое пространство украинской политической сферы, что позволит выделить конкретные стратегии в отношении будущего Донбасса, существующие сегодня в украинском политическом обществе. Вероятнее всего, эти стратегии будут занимать незначительное место в общем дискурсе. Это позволит сделать верифицированный вывод о том, что украинские элиты не имеют стратегического видения в отношении Донбасса, не воспринимают его реинтеграцию как реальную необходимость и используют феномен Донбасса преимущественно для поддержания уровня агрессии по отношению к России как “внешнему врагу” и собственной легитимации на основании образа».

Методология первой волны была простой: 1) Составить список тех блогеров, которых мы считаем формирующими общественное мнение, сделав это так, чтобы максимально избежать субъективности и пристрастий (каждый из участников проекта формирует ленту Facebook в соответствии со своими вкусами и, естественно, как это часто бывает в гуманитарных исследованиях, имеет соблазн подчинить общую канву собственным предпочтениям). 2) Скачать посты, написанные 176 блогерами в период с января по октябрь включительно. 3) Найти внутри них те, которые посвящены конкретно конфликту в Донбассе, предложив машине адекватный поисковый запрос. 4) Сформулировать собственное понимание возможных стратегий элит по отношению к Донбассу и, главное, выделить слова-маркеры, которые помогут выделить в общем дискурсе эти стратегии.

Мозговой штурм (поиск блогеров, определение возможных стратегий и их слов-маркеров, формирование поискового запроса), неформализованный экспертный опрос (верификация списка), big data (скачивание постов и поиск в этой БД необходимых) — таковы основные методы, которые планировалось использовать.

При помощи программного обеспечения «Семантический архив» был запущен мониторинг аккаунтов Facebook из сформированного списка — в базу были выгружены все записи, опубликованные с 1 января по 1 ноября 2016 г. В рамках пилотажного исследования этот период был выбран произвольно. Итоговый массив данных составил 88 536 записи — на этом база данных была окончательно сформирована.

Количество постов о Донбассе (5 200, то есть 5,87 %) в какой-то момент поставило под вопрос дальнейшее исследование — стоит ли продолжать, если это количество не соответствует нашим ожиданиям? Ведь мы были уверены, что тема конфликта стоит на первом месте в украинском сегменте соцсети.

В ходе «мозговых штурмов» было выделено четыре стратегии, которые, по нашему предположению, украинские лидеры общественного мнения могут использовать для определения будущего Донбасса: завоевание, экономическое вовлечение, отгораживание и заморозка конфликта. Попытка верификации через экспертные оценки не принесла успеха в том смысле, что ни один из опрошенных нами экспертов не смог предложить альтернативу этим четырем. Поэтому именно они были взяты за основу для распределения постов по рубрикам.

На первом этапе в качестве рабочей методологии машине было предложено ориентироваться на слова-маркеры для сортировки постов по выбранным четырем стратегиям. Логика выглядела следующим образом: если автор поста думает о завоевании Донбасса, он, скорее всего, будет использовать слова «боевики», «террористы» и т.д.; о заморозке конфликта — «отложить», «прекратить огонь» и т. д. (см. *Табл. 3*). Разумеется, сейчас подобная попытка жестко детерминировать множественные дискурсы широкого круга групп интересов во всем их лингвистическом разнообразии через привязку к крохотному количеству маркеров кажется наивной.

Таблица 3

Распределение слов-маркеров по стратегиям

Стратегии	Запрос	Кол-во постов
Завоевание	российско-террористические войска, боевики, террористы, вата, ватники, захватить, колорады, забрать, Россия-агрессор, сепаратисты, сепары	606
Заморозка	отложить, прекращение огня, заморозить, заморозка	34
Вовлечение	контроль, пенсии, децентрализация, реинтеграция, пацификация, прекращение огня, сограждане, гражданская война, шахты, заводы, переговоры	380
Отгораживание	отдать, забыть, Россия-агрессор, мертвая зона, стена	41

«Отгораживание» и «заморозка» с незначительной разницей остались в хвосте, «экономическое вовлечение» с явным отрывом вышло на второе место. С почти двукратным перевесом победила стратегия завоевания, что в итоге не имело никакого значения.

Мы подвергли выводы сомнениям по двум причинам, в итоге отказавшись от машинной обработки сформированного массива. Первая — количество полученных записей. Из 5 200 постов поисковые запросы со словами-маркерами суммарно выдали 1 061 пост, а значит, 4 139 постов (80 %) не содержали этих слов, то есть оказались исключенными из анализа. Это препятствие возможно было если не обойти, то максимально уменьшить его влияние, постоянно увеличивая число слов-маркеров, каждый раз захватывая всё больше и больше постов. Очевидно, что все 5 200 постов не стали бы объектом дискурс-анализа. Просто потому, что они собраны на основании наличия хотя бы одного из 14 синонимов, — к примеру, пост, состоящий из одного предложения «На Донбассе хорошая погода», тоже попадал в число 5 200, но, конечно, не имел никакого отношения к стратегиям.

Гораздо более весомым оказалось качественное препятствие, а не количественное. Здесь мы столкнулись с невозможностью отделить фактические высказывания людей от цитирований. Предположим, что условный украинский радикальный националист «Дмитрий Ярош» пишет пост: «Надо, наконец, победить российского агрессора: взять оружие в руки, сесть в танки и с Богом в душе зачистить украинский Донбасс от войск разлагающейся империи, сохранив за Украиной право на самостоятельное развитие, сохранив саму Украину. Слава Украине!». Прочитав его, условный сторонник федерализации и умиротворения Донбасса «Виктор Медведчук» делает репост с цитатой: «Есть те, кто призывает Украину к войне: “Надо, наконец, победить российского агрессора: взять оружие в руки, сесть в танки и с Богом в душе зачистить украинский Донбасс от войск разлагающейся империи”. Я категорически возражаю против такого человеконенавистнического подхода. Мы должны слышать всех сограждан, в том числе на востоке страны, настаивая на установлении режима прекращения огня».

Очевидно, что в одном и том же посте могут употребляться слова-маркеры из конфликтующих стратегий, либо могут быть использованы слова, подходящие к стратегии, с которой автор поста не согласен. Программное обеспечение не способно самостоятельно определить мнение автора поста — в этом случае оба поста с наличием слов-маркеров «агрессор», «оружие», «танк» уйдут в стратегию «завоевание» и в стратегию «заморозка» благодаря «прекращению огня» и «согражданам».

Несколько раз мы пробовали обсудить дальнейшую работу машины с разработчиками ПО, заверявшими нас, что нужно лишь предложить машине большее количество синонимов, с тем чтобы она лучше понимала наши стратегии. Увы, в этом споре технари проиграли гуманитариям, поскольку, как минимум, не удалось обойти препятствие цитирований. На данном этапе исследования мы присудили машине поражение и сели за вычитку 5 200 постов.

Собственно, именно это и свидетельствовало, что мы выкинули белый флаг, осознав крайнюю ограниченность метода Big Data, хотя разработчики наверняка с нами не согласятся, объясняя проигрыш метода нашей недостаточной работой по операционализации стратегий. Так или иначе, машина помогла нам найти в большом массиве данных 5 200 постов, которые и стали нашей конкретной базой данных (БД). Дальнейшая вычитка их «вручную» означала, что мы поняли ограничения автоматизированной работы.

В процессе прочтения записей украинских блогеров и поиска в них признаков одной из четырех стратегий мы ввели маркер «осмысленный пост» — так мы называли те записи, которые содержали хоть какой-то смысл по отношению к стратегии. Из 5 200 записей большинство содержало описание общей геополитической ситуации с точки зрения обывателя, обсуждение вопроса об аккредитации журналистов в ОРДЛО, новостные сводки с фронта, дебаты о субъектах, атаковавших «Боинг», и прочие не относящиеся к определению будущего региона размышления. Лишь в 327 постах (6,3 % от совокупности записей, посвященных региону) имелся хоть какой-то намек на то, как, по мнению автора, необходимо поступить с Донбассом. Для включения поста в рубрику с соответствующей стратегией было достаточно, например, фраз «Донбасс надо отгородить стеной», «пора уже завоевывать ОРДЛО». При отсутствии масштабных осмысленных стратегий — такие короткие фразы оказались для нас максимумом рефлексии, которая и была подвергнута сортировке.

117 постов относились к «заморозке» — их авторы настаивали на необходимости сохранения ситуации «ни мира, ни войны». К стратегии «завоевания» были отнесены 107 постов, об экономическом вовлечении говорилось в 69-ти, наконец, отгородиться стеной хотели авторы 34 постов. По отношению к общей БД в 88 536 постов число «осмысленных» составляло 0,37 %. Стоит ли говорить, что мы не стали вычислять весомость каждой из стратегий по причине ничтожности выделенной нами БД?

С другой стороны, на презентации доклада в ноябре 2016 г. по итогам первой волны уважаемые коллеги уверяли нас, что эти 327 постов внутри 5 200 постов и есть искомая нами БД. Следовательно, можно сделать вывод о практически полном отсутствии стратегии в отношении будущего Донбасса, поскольку лидеры общественного мнения Украины посвятили ему менее 0,5 % постов. В дальнейшем мы вынуждены были отказаться от поиска стратегий: машина не могла нам помочь из-за непонимания цитат, а мы сами убедились, что вычитка даже сотни постов — тяжелый труд. Учитывая, что после пилотной волны объем БД должен был вырасти многократно (во второй волне 5 200 постов сменились на 48 006), наше желание найти стратегии уступило рациональности.

* * *

Если украинский Facebook является важнейшим каналом коммуникации и если пороговое значение в 10 000 подписчиков и временные рамки исследования выбраны верно, то мы можем сделать следующий вывод: у украинской элиты нет доминирующей стратегии в отношении Донбасса. В России многим кажется, что на Украине только и говорят о возвращении Донбасса, но в реальности из всего того, что было написано лидерами общественного мнения в сети Facebook за неполный 2016 г., Донбассу посвящено менее 6 % постов. О конкретных же стратегиях (цели, задачи, методы решения) пока говорить не приходится. В публичном дискурсе украинских элит «Донбасс» представляет собой, скорее, территорию — геополитическую единицу, объект, на который направлены политические технологии или который участвует в юридических процедурах, — но не отколотый от «материнского государства» регион, в от-

ношении которого обсуждаются стратегии возвращения и будущее которого принципиально важно.

14,5 % упоминаний России в нашей выборке представляется незначительным для статуса «государства-агрессора». Вместе с тем специфика этих упоминаний должна быть дополнительно исследована. Означает ли это понижение степени россиецентричности украинского массового сознания? Или же украинский Facebook не разделяет магистрального отношения власти к пропаганде стереотипов «российской агрессии»? Поскольку отношение к России не было предметом данного исследования, объявить о переходе количества в качество в данном случае мы не можем.

На этом пилотажное исследование было завершено. Оно оказалось для нашей команды крайне полезным, поскольку позволило не только апробировать новый метод, понять его плюсы и ограничения, но и сформулировать новые вопросы. Например: Что представляет собой феномен «Донбасс» для лидеров украинского общественного мнения? Донбасс — это, прежде всего, территория или население? Является ли Донбасс частью украинского национального сообщества? Жители Донбасса, защищающие свои республики с оружием в руках, могут быть реинтегрированы в национальное сообщество? Есть ли у них право на политическое украинство при сохранении русской самоидентификации? Есть ли различия в восприятии украинскими элитами Донбасса, Крыма и особенно роли России в отношении обоих казусов? Для ответа на эти вопросы дискурс-анализ представляется многообещающим.

Ещё одно принципиально иное направление исследования с той же базой перспективно в отношении самих лидеров общественного мнения. Кто эти люди? Что они говорят, например, о России, формируя тренды в украинской политике?

Наконец, пилотажное исследование заставило нас изменить методологию.

Во-первых, необходимо было отказаться от критерия «формирование смыслов» при подготовке списка аккаунтов. Он был слишком размыт. Кроме того, наш субъективизм «лишал», например, звёзд эстрады права на политическое высказывание.

Во-вторых, мы должны были снижать проходной барьер, отказываясь от 10 000 подписчиков, поскольку за его пределами оказалось множество аккаунтов, которых мы были лишены при «пилотаже».

В-третьих, наше субъективное восприятие дискурса (в данном исследовании — представление о четырех стратегиях) тоже подверглось корректировке — при последующих волнах нам требовалось участие нескольких десятков коллег для экспертного опроса.

В-четвёртых, мы поняли, что ограничивать вокабуляр несколькими десятками слов в корне неверно. В дальнейшем мы его расширили до нескольких сотен слов.

В-пятых, был переработан поисковый запрос с увеличением количества слов от 14 до 110, что почти в два раза увеличило число постов, посвященных Донбассу, в базе данных.

Наконец, мы увидели значимые ограничения самой программы «Семантический архив». Она способна быть невероятно мелкой сетью, которая выловит из океана рыб необходимого размера. Но первоначальные надежды на то, что она проведет селекцию рыб и препарирует тех, на которых мы укажем, не оправдались.

ЛИТЕРАТУРА

1. Азаров А. А., Бродовская Е. В., Дмитриева О. В., Домбровская А. Ю., Фильченков А. А. Стратегии формирования установок протестного поведения в сети интернет: опыт применения киберметрического анализа (на примере «евромайдана», ноябрь 2013 г.) // Мониторинг общественного мнения. — 2014. — № 3. — С. 36–74.
2. Бродовская Е. В., Домбровская А. Ю., Синяков А. В. Стратегии использования социальных сетей в современной России: результаты многомерного шкалирования // Мониторинг общественного мнения. — 2016. — № 1. — С. 283–296.
3. Давыдов А. А. Системная социология: Social Computing. Институт социологии РАН. — URL: http://www.isras.ru/index.php?page_id=1016
4. Кобелева В. А., Бабенко П. С. Анализ взаимоотношений граждан России и Украины (на основе контент-анализа комментариев в сети Интернет) // Вестник Прикамского социального института. — 2017. — № 1. — С. 153–160.

5. Ляшенко И. В., Федюнина И. Э. Этнические прозвища русских в украинской и российской блогосферах // Научный результат. Серия: Вопросы теоретической и прикладной лингвистики. — 2017. — № 1. — С. 42–48.
6. Нечаева П. Е. Реакция российской элиты на присоединение Крыма к Российской Федерации // Аллея науки. — 2017. — Т. 2. — № 11. — С. 388–395.
7. Никопорец-Такигава Г. Ю. Информационная поддержка политических решений как политическая технология: из уроков недавнего прошлого // PolitBook. — 2016. — № 1. — С. 67–82.
8. Олещук П. Н. Влияние социальной сети Facebook на политическую мобилизацию // Гиля. Научный вестник. — 2014. — № 88. — С. 336–340.
9. Павлов П. В. Создание образа врага через посты и репосты политических медиановостей в социальной сети Facebook: семантический и прагматический аспекты воздействия // Известия Волгоградского государственного педагогического университета. — 2016. — № 4. — С. 90–97.
10. Савина С. С. Эмоциональная публицистика России и Украины в период 2013–2016 годов / Выпускная квалификационная работа по направлению «Журналистика». — СПб.: Санкт-Петербургский государственный университет, 2016. 76 с. — URL: <http://nauchkor.ru/uploads/documents/587d36335f1be77c40d58921.pdf>
11. Чернова Т. А., Слеповронская К. Ю. Историческая память в информационной войне. Как используются социальные сети в идеологическом противостоянии Украины и России // Философские науки. — 2015. — № 5. — С. 16–23.
12. Grassegger H. Ich habe nur gezeigt, dass es die Bombe gibt // Das Magazin. Aktualisiert am 10. Juni 2018. — URL: <https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/>
13. Leetaru K. Data mining methods for the content analyst: An introduction to the computational analysis of content. — Routledge, 2011. — 120 p.
14. Liu H., Salerno J., Young M. Social Computing and Behavioral Modeling. — Berlin: Springer, 2009. — 264 p.
15. Manovich L. The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics [2015]. — URL: <http://manovich.net/index.php/projects/cultural-analytics-social-computing>
16. Ronzhyn A. The use of Facebook and Twitter During the 2013–2014 Protests in Ukraine [July 2014]. — URL: https://www.researchgate.net/publication/268979057_The_Use_of_Facebook_and_Twitter_During_the_2013-2014_Protests_in_Ukraine
17. Wynn J. Digital sociology: Emergent technologies in the field and the classroom // Sociological Forum. — 2009. 24 (2). — P. 448–456.

А.А. ТОКАРЕВ

Big Data Applications in Social Media Research: The Experience of Unsuccessful Data Analysis of Ukrainian Facebook

Aleksei Tokarev, PhD (Political Science),
Senior Research Fellow, Center for Global Problems,
Institute for International Studies, MGIMO University.
E-mail: a.tokarev@inno.mgimo.ru

Summary. This article showcases a detailed description of the first stage of research on the discourse of Ukrainian opinion leaders on Facebook conducted by a team of researchers representing MGIMO University, Lomonosov Moscow State University, and Institute of Economy at the Russian Academy of Sciences. Convinced that it is Facebook that serves as the primary means of communication of politicians with the population in Ukraine, the team built a data base consisting of posts written over a 10-month period by 176 profiles belonging to the representatives of Ukrainian elites, and applied machine data analysis. The research question was the following: What strategies on the conflict in Donbas are verbalized by the Ukrainian elites? The author faced three challenges and limitations of machine data processing and analysis:

unsuccessful operationalization of terms; functional limitations of the Semantic Archive Platform, of which the author turned out to have unreasonably high expectations; lack of understanding of peculiarities of Big Data analysis. Nevertheless, it was the failure of this pilot research that helped raise crucial questions for further research, primarily on the criteria for shaping a data base and on formulating of the research questions for software. This experience turned to be essential for second and third stages of the research project that were completed a year and half after the project was launched. Hence the necessity to make public all the considerations on this research.

Keywords: big data, Facebook, Ukraine, Donbas conflict

REFERENCES

- Azarov A. A., Brodovskaya E. V., Dmitrieva O. V., Dombrovskaya A. Y., Filchenkov A. A. Strategii formirovaniya ustanovok protestnogo povedeniya v seti internet: opyt primeneniya kibermetricheskogo analiza (na primere evromajdana, noyabr' 2013 g.) [Strategies for the Protest Behavior Formation in the Internet: The Experience of Using Cybermetric Analysis (Using the Example of EuroMaidan, November 2013)]. *Monitoring obshchestvennogo mneniya*. 2014. No. 3. P. 36–74.
- Brodovskaya E. V., Dombrovskaya A. Y., Sinyakov A. V. Strategii ispol'zovaniya social'nyh setej v sovremennoj Rossii: rezul'taty mnogomernogo shkalirovaniya [Strategies of Using Social Networks in Modern Russia: The Results of Multidimensional Scaling]. *Monitoring obshchestvennogo mneniya*. 2016. No. 1. P. 283–296.
- Chernova T. A., Slepovronskaya K. Yu. Istoricheskaya pamyat' v informacionnoj vojne. Kak ispol'zuyutsya social'nye seti v ideologicheskom protivostoyanii Ukrainy i Rossii [Historical Memory in the Information War. How Social Networks are Used in the Ideological Confrontation between Ukraine and Russia]. *Filosofskiye nauki*. 2015. No. 5. P. 16–23.
- Davydov A. A. Sistemnaya sociologiya: Social Computing [System Sociology: Social Computing]. *Institute of Sociology of the Russian Academy of Sciences*. — URL: http://www.isras.ru/index.php?page_id=1016&printmode.
- Grassegger H. Ich habe nur gezeigt, dass es die Bombe gibt // *Das Magazin*. Aktualisiert am 10. Juni 2018. — URL: <https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/>
- Kobeleva V. A., Babenko P. S. Analiz vzaimootnoshenij grazhdan Rossii i Ukrainy (na osnove kontent-analiza kommentarijev v seti Internet) [Analysis of the Relationship between Citizens of Russia and Ukraine (Based on Content Analysis of the Internet Comments)]. *Vestnik Prikamskogo sotsial'nogo instituta*. 2017. No. 1. P. 153–160.
- Leetaru K. *Data Mining Methods for the Content Analyst: An Introduction to the Computational Analysis of Content*. Routledge, 2011. 120 p.
- Liu H., Salerno J., Young M. *Social Computing and Behavioral Modeling*. Berlin: Springer, 2009. 264 p.
- Olyashenko I. V., Fedyunina I. E. Etnicheskie prozvizhcha russkikh v ukrainskoj i rossijskoj blogosferah [Ethnic Nicknames of Russians in the Ukrainian and Russian Blogospheres]. *Nauchnyj rezul'tat. Voprosy teoreticheskoy i prikladnoj lingvistiki*. 2017. No. 1. P. 42–48.
- Manovich L. The Science of Culture? *Social Computing, Digital Humanities and Cultural Analytics* [2015]. URL: <http://manovich.net/index.php/projects/cultural-analytics-social-computing>
- Nechaeva P. E. Reaciya rossijskoj ehlyty na prisoedinenie Kryma k Rossijskoj Federacii [The Reaction of the Russian Elite to the Crimea Reintegration]. *Alleya nauki*. 2017. Vol. 2. No. 11. P. 388–395.
- Nikiporets-Takigawa G. Y. Informacionnaya podderzhka politicheskikh reshenij kak politicheskaya tekhnologiya: iz urokov nedavnego proshlogo [Information Support of Political Decisions as a Political Technology: From the Lessons of the Recent Past]. *PolitBook*. 2016. No. 1. P. 67–82.
- Oleshchuk P. N. Vliyanie social'noj seti Facebook na politicheskuyu mobilizaciju [The Influence of Facebook on Political Mobilization]. *Gileya. Nauchnyy vestnik*. 2014. No. 88. P. 336–340.
- Pavlov P. V. Sozdanie obraza vraga cherez posty i reposty politicheskikh medianovostej v social'noj seti Facebook: semanticheskij i pragmaticheskij aspekty vozdejstviya [Creation of the Enemy Image by Means of Posts and Reposts of Political Media News in Facebook: The Semantic and Pragmatic Aspects of the Impact]. *Izvestiya Volgogradskogo gosudarstvennogo pedagogicheskogo universiteta*. 2016. No. 4. P. 90–97.
- Ronzhyh A. *The use of Facebook and Twitter During the 2013–2014 Protests in Ukraine* [July 2014]. — URL: https://www.researchgate.net/publication/268979057_The_Use_of_Facebook_and_Twitter_During_the_2013-2014_Protests_in_Ukraine
- Savina S. S. *Emocional'naya publicistika Rossii i Ukrainy v period 2013-2016 godov. Vypusknaya kvalifikatsionnaya rabota*

po napravleniyu "Zhurnalistika". [Emotional Publicism of Russia and Ukraine in 2013–2016. Graduation Qualification Thesis in Journalism]. St. Petersburg State University, 2016. 76 p. — URL: <http://nauchkor.ru/uploads/documents/587d36335f1be77c40d58921.pdf>

Wynn J. Digital sociology: Emergent Technologies in the Field and the Classroom. *Sociological Forum*. 2009. No. 24 (2). P. 448–456.