

6. АНАЛИЗ НЕСТРУКТУРИРОВАННОЙ ИНФОРМАЦИИ: ЗАПРОС НА МЕТОДОЛОГИЮ

Сетевой анализ постсоветского информационного пространства: перспективы разработки методологии

Градосельская Галина Витальевна, *НИУ ВШЭ*

1. Политические и информационные процессы: взаимодополнение

В настоящее время интенсивность информационного воздействия на сознание человека резко возросло. Имеется в виду, как усиление влияния традиционных информационных факторов, таких как СМИ, так и расширение способов информационного воздействия, их изменение, например, переход к интерактивному телевидению, и, конечно, коммуникация через Интернет. При этом возрастает посредническая роль текста как проводника социального воздействия: через СМИ реализуются манипулятивные технологии (причем на массовом уровне и оказывают влияние на общую социальную жизнь); изменяется образ общения, во многом переходя в виртуальное пространство и за счет использования новых технологий...

Политические акторы и их деятельность погружена одновременно в экономическую, политическую и информационную плоскости. Информационное пространство, выполняя манипулятивные функции в отношении населения разных стран, вносит существенный вклад в легитимацию принимаемых политических решений. Помимо привычных СМИ политическое проектирование все более осваивает интерактивную Интернет-среду, гибко улавливая формирующиеся настроения населения, и одновременно незаметно ими манипулирует.

На рис. 1 показаны уровни конструирования информационного пространства от речевых маркеров до плоскости коммуникативного взаимодействия авторов и манипулятивного воздействия коллективных акторов.

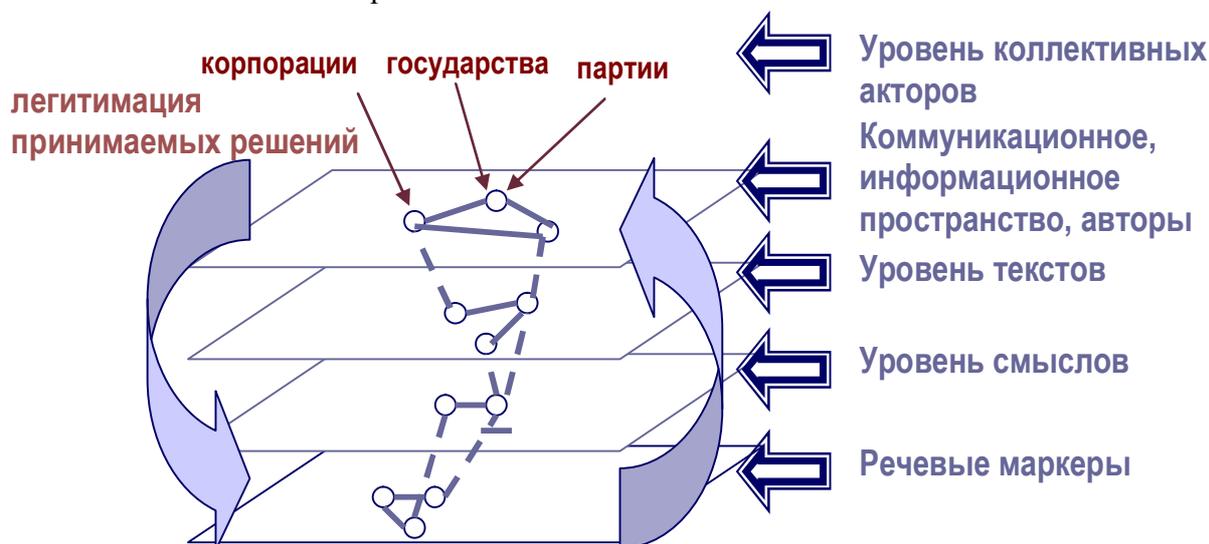


Рис. 1. Уровни анализа неструктурированной информации

Актуальнейшим направлением развития сетевой методологии является анализ информационного пространства, особенно в свете расширяющегося влияния Интернет-среды, которое в значительной степени «зашумлено». Назрела потребность в инструменте, который позволял бы направленно анализировать гибко меняющиеся веяния в Интернет-пространстве, группировать акторов (участников информационной игры), анализировать расстановку политических сил и их ходов. На данный момент PR и информационные службы государственных

ных структур и крупных корпораций России до сих пор пользуются устаревшими методами, что приводит к существенному запаздыванию и проигрышу в информационных войнах.

Особое направление текстового анализа представляет структурирование текстов, взаимного позиционирования акторов, а так же контекстного сопровождения коммуникативного действия в информационном пространстве. Усилия многих IT-компаний направлены на автоматизацию исследований информационного пространства. Представляется, что скромность результатов связана с отсутствием содержательно-ориентированной методологии, четкого формулирования одновременно содержательных и типизируемых задач. Сильно недооценено взаимодействие между «предметником» (специалистом, понимающим содержательную цель анализа текстов), «методологом» и «алгоритмистом/программистом». На данном этапе представляется необходимым обсуждение накопившихся проблем анализа неструктурированной информации со стороны исследователей, занимающихся практическими проблемами и пересекающимися прикладными задачами, находящимися на стыке наук.

2. Перспективы развития методологии анализа неструктурированной информации

Социальные процессы в современном обществе претерпевают критические изменения в своих проявлениях, формах воздействия и возможностях их научной рефлексии. Можно сказать, что реальные социальные взаимодействия и на макро-, и на микроуровне все более переносятся в информационно-коммуникативную плоскость. Роль текста как посредника социального окружения, благодаря которому происходит включение индивида в социальную реальность, резко возрастает.

Современные формы социально-лингвистических процессов и их текстовое выражение исследуют разные науки: социология, социолингвистика, математическая лингвистика, прикладная математика, системы автоматизированного проектирования, политология, психология, маркетинг и др. Все эти разноплановые гуманитарные и технические дисциплины выбрали один и тот же объект — социально-лингвистическое сопровождение коммуникативной активности человека. Совмещение традиционных подходов из разных наук на теоретическом, методологическом и прикладном уровнях повышает актуальность разработки пограничного инструментария, позволяющего оперативно проверять социально-коммуникативные и социально-лингвистические гипотезы, а так же расширять содержательное поле анализа.

Методы исследований неструктурированной информации разделены на количественные и качественные подходы, и наблюдается дифференциация этих подходов по разным дисциплинам. Количественные методы анализа неструктурированной информации развивались довольно обособленно, больше в технических дисциплинах — искусственный интеллект, нейросетевое моделирование, лингвистическое обеспечение систем автоматизированного проектирования и программирования и т.п. Качественные методы, концентрирующиеся на содержательной концептуализации и операционализации происходящих явлений, развивались в социолингвистике, социологии, журналистике и маркетинге.

Разные науки решали сходные содержательные объективные проблемы параллельно, в собственных методологических традициях, не зная, что делается в других науках. Настал критический момент для обсуждения общего предметного поля, и более экономного (и вместе с тем более эффективного) распределения научных сил, чтобы концентрированно решать объективные задачи.

Уровень вопросов, рассматриваемых при анализе неструктурированной информации, варьирует от принятия управленческих решений (включая государственные и/или корпоративные) до формализации задач и разработки математических алгоритмов или решения лингвистических проблем. Но социальные науки оказались не готовы в достаточной степени к концептуализации новых научно-прикладных задач, возникающих в современном социально-информационном пространстве.

Необходимо понимание и операционализация новых методологических понятий (объекта исследования, постановки задач и др.), а так же инвентаризация уже разработанных методов, активно используемых в разных науках при исследовании неструктурированной информации. В любом случае актуальнейшим вопросом является разработка методологии анализа

неструктурированной информации как последовательной цепочки исследовательских процедур: от постановки и формализации задач до разработки типовых решений. Также на современные задачи анализа социальной информации оказывают значительное влияние новые информационно-коммуникативные технологии: динамические (электронные СМИ) и структурные (создание и управление социальными сетями и т.д.). Несмотря на динамичное развитие объекта исследования, инструментарий анализа, в том числе применение прикладных программ, развивается скорее в догоняющем темпе.

Актуальными представляются следующие **направления содержательного развития** анализа неструктурированной информации: понимание объекта и предмета исследований; перспективы развития методологии анализа неструктурированной информации; возможности сочетания анализа неструктурированной информации с математическими методами анализа данных. Также существуют **технологические направления** потенциального развития неструктурированной информации: возможности дальнейшего развития имеющихся пакетов обработки неструктурированной информации; перспективы приложения интеллектуального анализа данных (Data Mining) к исследованию неструктурированной информации. Особенный интерес представляют **направления структурного анализа** информационной среды: особенности анализа неструктурированной информации в Интернет-среде, сервисах и т.д.; анализ социальных сетей и структурирование информации.

Прорыву в развитии методов мешает отсутствие выстроенной «с конца» методологической цепочки, объединяющей по содержательному основанию все остальные процедуры. Приоритетом является практическая постановка задачи в виде ожидаемых результатов и четкого понимания целей разрабатываемой методологии (рис. 2).

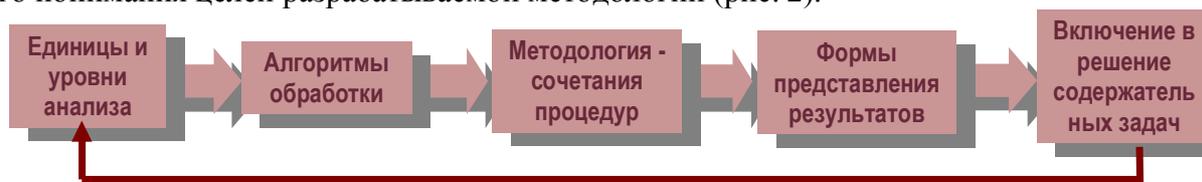


Рис. 2. Методологическая цепочка процедур для анализа неструктурированной информации

Реализация каждой из процедур также может производиться разными алгоритмами. Поэтому представленная принципиальная методологическая цепочка ни в коем случае не является ограничением для количества алгоритмов и подходов, реализуемых представителями разных наук. Повышается значимость их итеративного взаимодействия. В перспективе развития методологии на каждом этапе обработки неструктурированной информации у аналитика должна быть возможность выбора алгоритма и их комбинации для решения целевой задачи. Возможность реализации полной методологической цепочки показан ниже на примере разработанной автором методологии структурирования текстов при помощи зерновых концептов.

3. Методология структурирования текстов при помощи зерновых концептов: на примере концептуальной оппозиции «Россия–США»

Проблема структурирования текста (в формате небольшой статьи, в которой обычно предоставляются экспертные мнения) имеет длительную историю попыток ее алгоритмического решения. При этом аналитику приходится постоянно выбирать условия реализации этой процедуры. Предварительное структурирование текста представляется актуальной задачей, поскольку структура позволяет совершить принципиальный переход от линейно организованного текста к системе связанных элементов текста. Кроме того, по структуре можно определять скрытые смыслы, заложенные авторами текста (не секрет, что большинство экспертных изданий являются собственностью крупных корпораций или государственных структур, поэтому в публикуемых статьях должны соблюдаться интересы собственников издания).

Преыдушие исследователи старались максимально подробно воспроизвести структуру текста по разным основаниям (например, через синтаксический разбор текста и т.п.). При

этом структура текста становилась самоцелью, без понимания финального приложения разрабатываемых методов в финальной аналитике. Предполагалось, что у текста есть некоторая абсолютная структура (инвариантная к содержательной цели анализа текста), и ее воспроизводство может помочь в анализе прикладных проблем. С точки зрения практического анализа, можно указать два ограничения такого подхода. Во-первых, даже если такой алгоритм для русского языка и будет разработан, его автоматизация потребует значительных машинных ресурсов. Во-вторых, в случае неограниченных текстовых массивов (Интернет) анализ структуры текста (даже если это получится) становится бессмысленным и невозможным физически.

Анализируя большой корпус экспертных публикаций в режиме информационного потока в течение нескольких лет, мы заметили, что формат небольших экспертных статей имеет некоторые общие законы построения: концентрированность на одной теме, некий сценарный план развития, который от «затравки» (информационного повода) приводит читателя к заранее заложенным выводам. Все это обеспечивается очень четкими, отточенными предложениями, где речевые маркеры весьма четко сопрягаются и связывают между собой основные тематические концепты. Подобного подхода придерживаются при так называемом «осевом кодировании» в обоснованной теории. Идея весьма очевидна, а алгоритмов автоматизации может быть множество, что и составляет открытое поле для методологических экспериментов.

В большинстве случаев экспертные статьи вскрывают некоторый конфликт, оппозицию сторон, а затем отдается предпочтение (советы) одной из них, приглашая читателя присоединиться к выводам автора. В простейшем случае оппозиция в тексте только одна: «Россия vs. США», «Россия vs. ВТО», «Опель vs. Автоваз» и т.д. Противопоставляемые «сущности» выражаются конкретными речевыми маркерами или концептами. Здесь, с точки зрения аналитической ценности, важно понять, каким образом автор текста связывает эти оппозиции, посредством каких речевых маркеров, или их групп (можно провести аналогию с дугами напряжения, связывающими электрические полюса). Тогда и построение структуры становится гораздо проще: отпадает необходимость в большей части связей (не существенных для аналитической цели) и в некоторой части речевых маркеров (что также весьма облегчает построение итоговой структуры).

Безусловно, наш алгоритм структурирования экспертных текстов методом зерновых концептов так же имеет ограничения, например, в применении к анализу очень больших текстов или неэкспертных текстов (глубинных интервью или фокус-групп). Однако мы и не претендуем на всеобщность. Для структурирования транскриптов «живой речи» имеет смысл разработать другой, наиболее подходящий алгоритм.

Алгоритм структурирования экспертных текстов при помощи зерновых концептов можно разбить на следующие этапы:

- задаем зерновые концепты статьи — основные смысловые термины;
- проводим предварительный дискурс-анализ, где основными «центрами напряжения» становятся зерновые концепты;
- остальные концепты добавляются только в том случае, если они обеспечивают связь (посредничество) между зерновыми концептами;
- укрупнение посредничающих концептов в «смысловые гнезда»;
- впоследствии можно провести автоматизацию алгоритма (в нашем случае автоматизация была проведена при помощи программы MatLab).

В примере мы опускаем ряд технических, но, тем не менее, важных этапов построения структуры: нормализацию исходного текста, очистку от стоп-слов, построение матрицы, а так же алгоритм поиска прямых путей между зерновыми концептами для дальнейшего преобразования в двухполюсный граф, поскольку иначе наша статья может увеличиться в три раза.

В качестве примера для построения структуры текста методом зерновых концептов была выбрана статья «Московский саммит: нажимая не на те кнопки» («The Guardian», Великобритания, 06.07.2009). Эта небольшая статья, посвященная будущей встрече Б. Обамы и Д. Медведева, реализует принцип оппозиции и содержит несколько скрытых смыслов, нега-

тивных для России. Статья является хорошим примером того, как в информпространство запускается журналистская метафора, емко характеризующая происходящие социально-политические процессы. В этой статье содержится одно из первых упоминаний термина «перезагрузка», подразумевающего обновление отношений между Россией и США. При этом основной текст статьи полностью противоречит декларируемой «перезагрузке», поскольку представляет историю российско-американских отношений как череду сплошных неудач. Ключевыми концептами в этом случае будут «Россия» и «США». Итоговая схема показана на рис. 3.

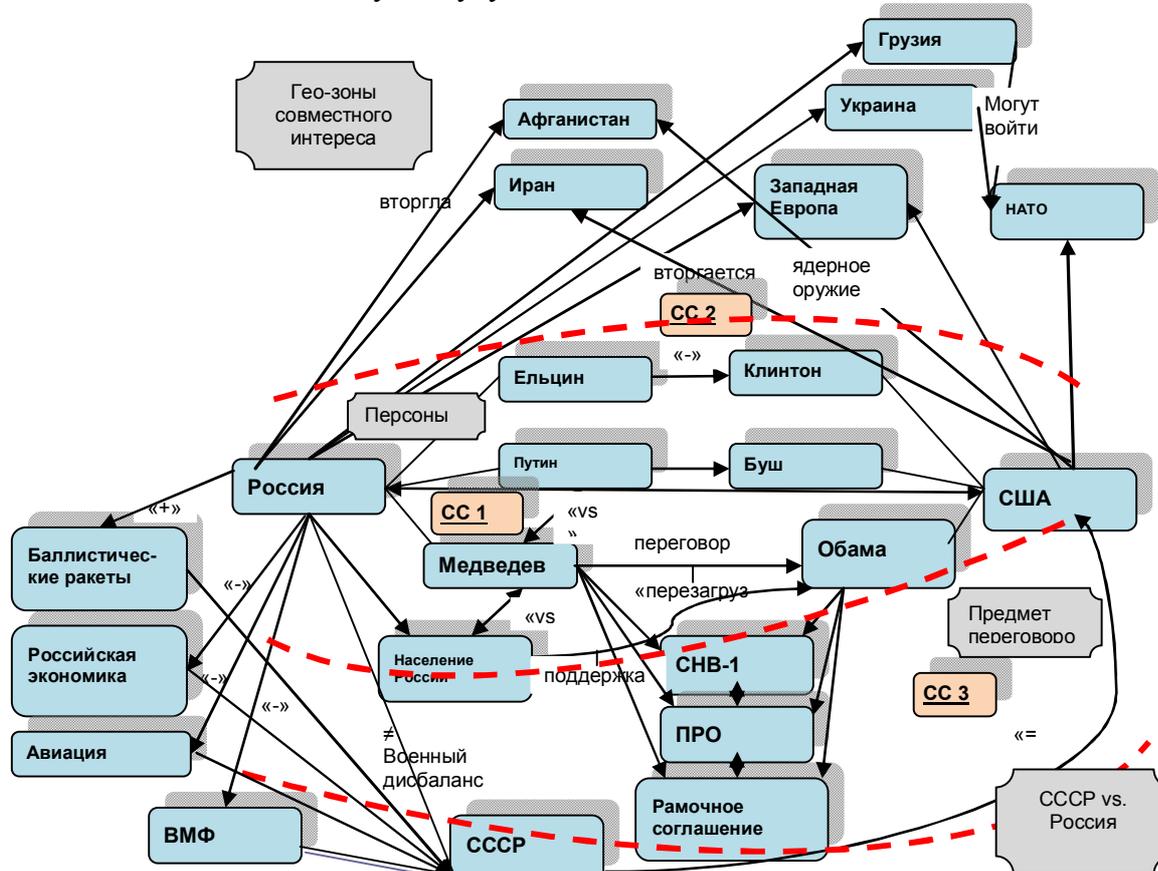


Рис. 3. Дискурс-анализ статьи «Московский саммит: нажимая не на те кнопки» («The Guardian», Великобритания, 06.07.2009) по принципу зерновых концептов

Статья состоит из 4 абзацев, в каждом из которых выстраивалась цепочка посредничающих смыслов, связывающих концепты Россия и США. Они касались текущей геополитической ситуации и роли ключевых государств Ближнего Востока, Западной Европы и постсоветского пространства (Украины и Грузии). Также рассматривалась история предыдущих контактов между главами двух стран (Ельцин–Клинтон, Путин–Буш). В последнем абзаце на ряде противопоставлений современной России и СССР (в экономике, вооружении, демографической ситуации) проводилась идея об утрате Россией потенциала и ресурсов, которые были в период СССР.

Фактически, каждый абзац реализовывал скрытый смысл (СС на рис. 3):

Скрытый смысл 1 (СС1): «Медведев не репрезентирует настроения своих граждан».

Скрытый смысл 2 (СС2): «Все контакты между Россией и США на личном уровне президентов в прошлом терпели крах».

Скрытый смысл 3 (СС3): = резюме статьи: «Медведев — не тот человек, с которым нужно вести переговоры и который отвечает за действия государства. Переговоры нужно вести как государство с государством».

Самый ценный результат этого алгоритма заключается в том, что его удалось запрограммировать при помощи пакета MatLab и получить «сгустки» речевых маркеров, каждый из которых воспроизводил один из ключевых концептов, показанных на рис. 3. В дальнейшем, ка-

кие-то концепты из данной статьи могут стать зерновыми для других экспертных статей; так наша структура будет расширяться методом «снежного кома».

4. Анализ информационного пространства ближнего и дальнего зарубежья: примеры

Ниже показаны весьма сокращенные примеры постановки содержательных задач и возможности структурно-сетевому анализу информационного пространства по результатам собственных исследований автора на международном уровне (в т.ч. стран СНГ).

4.1. Сегментация информационного поля Казахстана в период передела информационного пространства (2008)

Рассмотрен период политической нестабильности Казахстана, когда возникла угроза смещения Н. Назарбаева со стороны его первого зятя Р.Алиева. На основе анализа корпуса аналитических статей (более 40 за 2007 г.) рассмотрен процесс формирования оппозиции по сетевому принципу. Было показано, как Р. Алиевым на основе кланово-родственных и дружеских связей выстраивалась сеть влияния. Начиналось с расстановки «своих людей» в силовых структурах, потом с помощью влияния в силовых структурах в сеть были включены финансовые структуры. На следующем шаге одновременно были сделаны два шага: перехват контроля над информационным пространством и создание собственной политической партии, оппозиционной действующему правительству.

Был разработан алгоритм, позволяющий выделять из линейно организованного текста действующих политических акторов и организации, которые так же влияли на организацию информационно-политического пространства страны.

4.2. Методология построения межстрановой когнитивной карты в электронных СМИ (на примере обсуждения проблемы вступления РФ в ВТО)

Проблема вступления России в ВТО давно является предметом политических торгов и манипуляций и в России, и за рубежом. В нашем исследовании (которое было выполнено совместно с Д. Проснянук) было проанализировано более 100 статей за 2008–2010 гг. в России и за рубежом, посвященных возможности вступления России в ВТО.

На рис. 4 показано, что распределение реальных торговых потоков (товарооборота) между Россией и другими странами отличается от распределения объемов публикаций зарубежных страны о возможности вступления России в ВТО. Больше всего за период 2008–2010 гг. о вступлении России в ВТО писали в Великобритании (занимает лишь 12-е место по торговому обороту с нашей страной), в США (9-е место) и Бельгии (17-е место). При чем США и Великобритания в основном публиковали негативные статьи о перспективах вступления России в ВТО.

Анализ аргументов зарубежных стран и российской стороны показал существенную коммуникативную асимметрию: аргументы, которые высказывает Россия, никак не учитываются западными странами и наоборот. В результате была построена межстрановая когнитивная карта, которая показала структуру поддержки и противодействия вступлению России в ВТО со стороны других стран.

4.3. Механизмы и направления мифологизации образа России на постсоветском Интернет-пространстве (на примере Украинских форумов)

Отличие этого примера от предыдущих в том, что в качестве объекта выступали не экспертные тексты, а живые реплики обычных пользователей Интернета, высказывающихся на украинских политических форумах. Совместно с К. Маркиной были исследованы более 100 форумов с политическими темами, за 2009 г.

В процессе работы с постсоветским информационным пространством (прежде всего казахским и украинским), мы убедились, что оно явно подвержено манипуляциям со стороны «третьих лиц» в направлении, негативном для имиджа России. Проявляется это, прежде всего, в сходстве приемов, демонстрируемых в казахской и украинской блогосфере: подмене истории, навешивании ярлыков и т.д.

В случае с украинскими форумами, побочным результатом исследования стало выявленная принадлежность половины украинских форумов госдепу США.

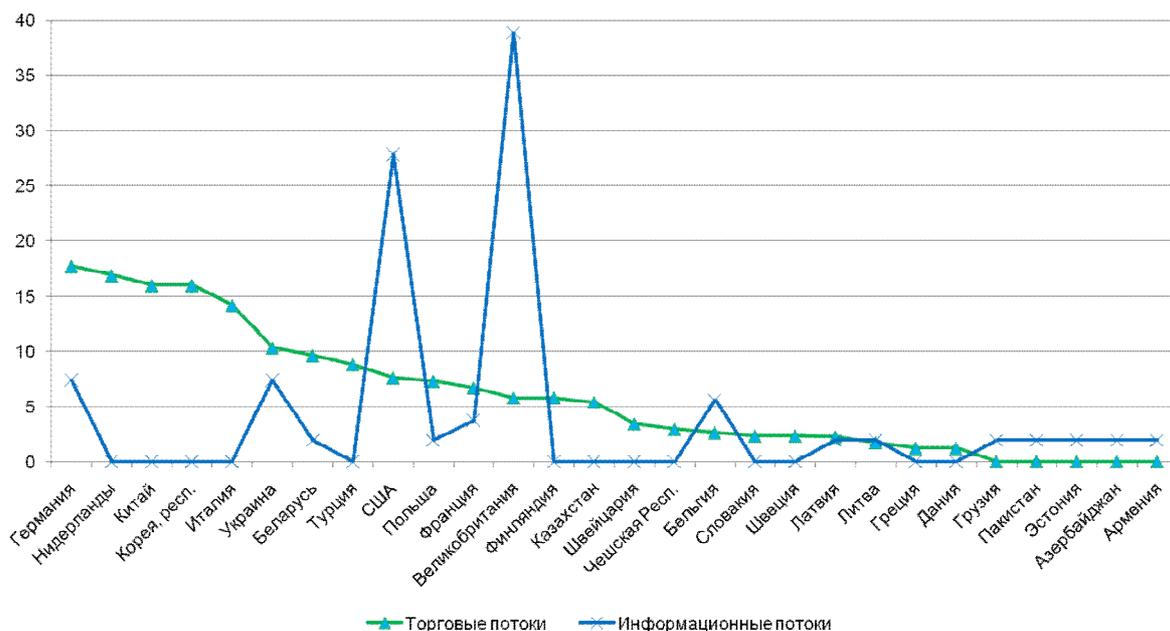


Рис. 4. Несовпадение торговых и информационных потоков: торговый оборот между Россией и другими странами, публикации этих стран о вступлении России в ВТО, (% от объема потоков, суммарного всех стран).

Особую методологическую ценность данному исследованию придает использование методов многомерной статистики (многомерного шкалирования и кластерного анализа) к анализу группировки смыслов, концептов и аргументов авторов в создании негативного (по большей части) образа России в информационном пространстве Украины.

Заключение

В современном информационном пространстве происходят бурные процессы. Объемы анализируемой информации растут, растет и технологическая машина производства текстов и смыслов. Особенно уязвимыми становятся коллективные акторы, репрезентирующие большие распределенные структуры (государства и корпорации). Понимание того, что происходит в окружающем эти структуры информационном поле, какие смыслы представляют угрозу, а какие работают на благо коллективных акторов, помогает ориентироваться и планировать собственные стратегические ходы. В самой России, в ближнем зарубежье, а так же странах англо-саксонского мира стали общеприняты стандарты дискредитации России, ее исторического прошлого и современной политики. Структурно-методологический анализ поможет широким слоям экспертного и научного сообщества понять это, прежде всего, через предоставление иллюстративных и проективных исследований.

Если перейти на индивидуальный уровень, то при общем падении культуры населения (и возросшей подверженности манипуляциям) повышается их включенность в информационные системы, что многократно усиливает риск неконтролируемых конфликтных всплесков (как произошло на Ближнем Востоке в 2011 г., или в Киргизии в 2010 г.). Современное общество все более атомизируется, возникают новые основания для временной солидарности, которая может основываться на форме (например, протестной), а не на содержании или традициях. Фактически текст становится более важной формой проектирования социального действия, чем экономические основания. Поэтому запрос на методологию изучения информационного пространства и его структурирования становится как никогда актуальным.

Основные проблемы развития методологии структурирования текстов и информационного пространства заключаются в отсутствии методологической цепочки взаимосвязанных процедур. Необходимы четко сформулированные содержательные задачи анализа текстов; понимание содержательных и ресурсных ограничений разрабатываемых методов; возможности под-

бора наиболее подходящих математических алгоритмов для обработки информации и адекватных форм для представления результатов. Для развития анализа неструктурированной информации необходимы междисциплинарный взгляд и поэтапное, итеративное обсуждение с представителями разных профессиональных областей. Одна из первых попыток такого комплексного междисциплинарного обсуждения и были предприняты в рамках этой секции.

Социальные технологии изучения коммуникационных процессов в современной исследовательской практике

Адамьянц Тамара Завеновна,
Институт социологии РАН

Для анализа социальных последствий, инициированных информационной (шире — социокультурной средой), необходимы данные об особенностях отражения отдельных произведений, а также и в целом социокультурной (информационной) среды, в сознании и «картинах мира» представителей разных групп аудитории. Решение такой задачи подразумевает комплекс взаимосвязанных методов и процедур, позволяющих переход от теории и деклараций к практическому уровню, причем посредством доказательных и воспроизводимых данных. Речь, следовательно, идет о *социально-диагностических и социально-проектных технологиях*, основанных на междисциплинарном анализе: попытки решения отдельных, узких задач, в отрыве от других, взаимосвязанных с ними, как показывает опыт, малопродуктивны.

Методологическим и методическим аппаратом, адекватным для решения поставленных задач, оказывается комплексная междисциплинарная семиосоциопсихологическая концепция, позволяющая глубинное понимание общих, единых для всех форм общения и взаимодействия людей механизмов функционирования социальной коммуникации в обществе¹. Это принципиально новое поле анализа коммуникационных процессов, где во главе угла оказываются *интенциональные (мотивационно-целевые)* первопричины общения, а основными целями коммуникации оказываются *адекватное понимание и взаимопонимание*, без чего невозможно конструктивное общение и взаимодействие людей.

В отличие от лингвистики и психолингвистики, изучающих *значения* отдельных слов, фраз или дискурсов (такой анализ также необходим), предметом изучения в семиосоциопсихологии оказываются целостные, завершенные коммуникативные акты, реализованные в любой знаковой системе или в их комплексе (фильмы, стихотворения, телепередачи, стихотворения, компьютерные игры, романы, устные выступления и т.д.), а искомой точкой отсчета (той, по отношению к которой происходит понимание) — авторская *интенциональность*, или «равнодействующая мотивов и целей общения и взаимодействия людей»². Здесь, в интенциональности, кроется основной *смысл* (смысловая доминанта) коммуникативного акта: это то *самое главное*, что хотел сказать, передать, выразить автор; это тот искомый результат, к которому он стремился, причем и на уровне осознанных целей, и на уровне не всегда осознаваемых мотивов.

Процедура поиска интенциональности коммуникативных актов операционализирована, и потому результат является доказательным и воспроизводимым. Об этой процедуре опубликовано большое количество научных работ, поэтому основной акцент настоящей статьи сосредоточен на комплексных технологиях сочетания данной процедуры с традиционными социологическими и социально-психологическими методами и подходами, что позволяет выходить на уровень социально-диагностических выводов и социально-проектных рекомендаций. И все же несколько фраз о методологических основаниях, позволяющих данную процедуру.

¹ Дридзе Т.М. Текстовая деятельность в структуре социальной коммуникации. М. Наука, 1984.

² Дридзе Т.М. Две новые парадигмы для социального познания и социальной практики // Социальная коммуникация и социальное управление в эквантропоцентрической и семиосоциопсихологической парадигмах. Книга 1. М. ИС РАН, 2000. С. 16.