

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: взаимодействие

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.

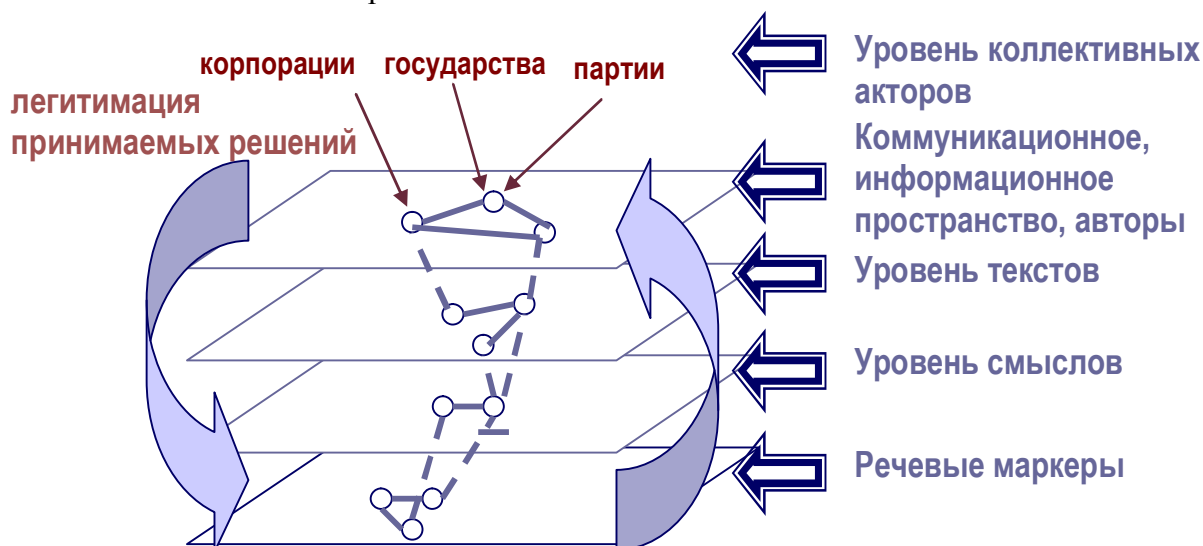


Рис. 1. Уровни анализа неструктурированной информации

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. Особый 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).

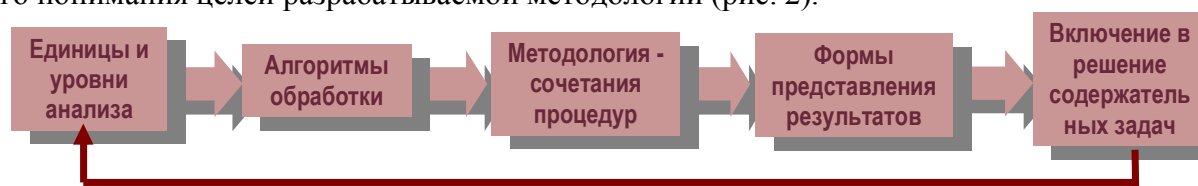


Рис. 2. Методологическая цепочка процедур для анализа неструктурированной информации

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).

Предыдущие 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. Зерновыми 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

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

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

Самый ценный результат этого алгоритма заключается в том, что его удалось программно реализовать при помощи пакета MatLab и получить «сгустки» речевых маркеров, каждый из которых воспроизводит 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-сетевого анализа 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. Просянюк) 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 товарообороту 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.

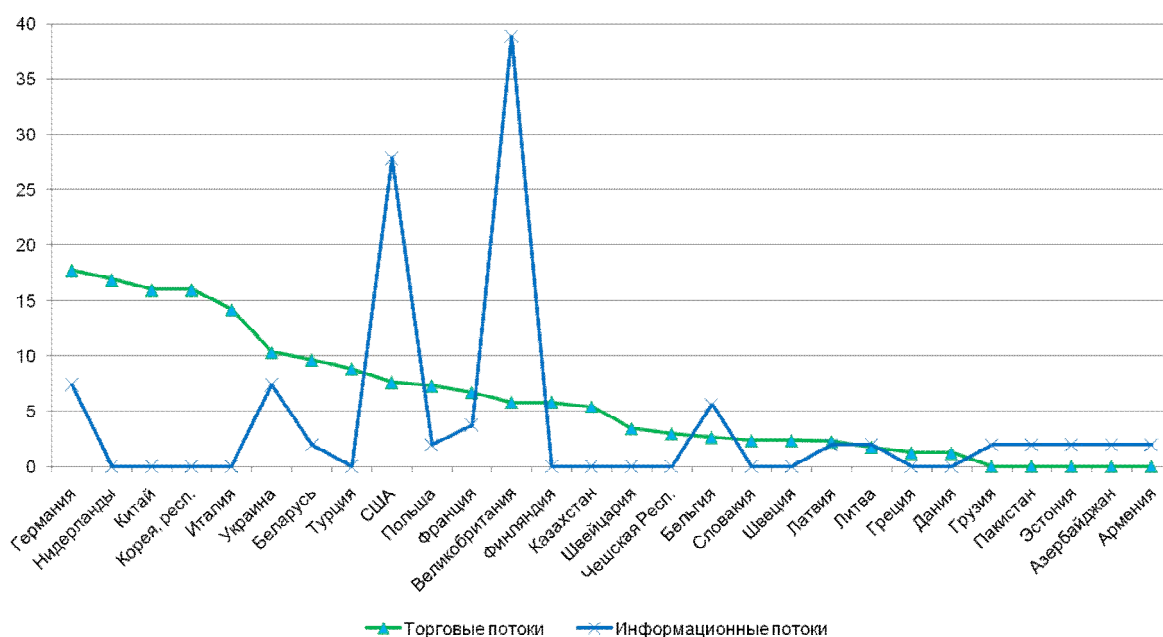


Рис. 4. Несовпадение торговых и информационных потоков: торговый оборот между Россией и другими странами, публикации этих стран о вступлении России в ВТО, (% от объема потоков, суммарного всех стран).

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 экوانтропоцентрической и семиосоциопсихологической парадигмах. Книга 1. М. ИС РАН, 2000. С. 16.

Любой целостный, завершённый коммуникативный акт, согласно одному из основных положений семиосоциопсихологической концепции, можно условно представить как иерархически организованную структуру коммуникативно-познавательных программ, ориентированных на интенцию, подчинённых ей¹. Иными словами, в любом целостном, завершённом коммуникативном акте всегда есть более и менее главные по отношению к основной интенциональности уровни коммуникативно-познавательных программ (тезисы, аргументы, иллюстрации, фоны, фоны к фонам и т.д.), имеющие, в свою очередь иерархические отношения между собой (см. рис. 1). На искусстве выделения таких структур и строится основной исследовательский метод семиосоциопсихологии — метод интенционального (мотивационно-целевого) анализа.

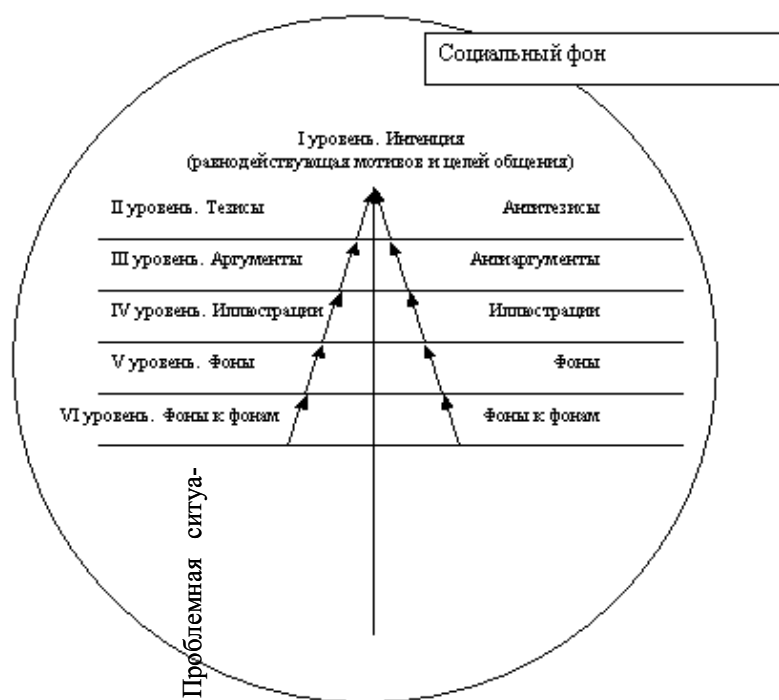


Рис.1. Типовая интенциональная (мотивационно-целевая) структура целостного, завершённого коммуникативного акта

Коммуникативные акты практически всегда, даже если они однотипные, различаются «набором» элементов внутри структур и способами реализации интенциональности. Зримо «увидеть» эти структуры и, таким образом, доказательно определить смысл исследуемого произведения (материала) позволяет интенциональный (мотивационно-целевой) анализ. Такая работа проводится при социальной диагностике коммуникационных процессов и отдельных произведений; при этом для каждого произведения, ставшего предметом изучения, составляется отдельная структура. Постигание общих принципов интенционального (мотивационно-целевого) анализа позволяет не только выявлять латентные механизмы в коммуникации, но и развивает навыки понимания (*коммуникативные навыки*): человек учится различать в воспринятом произведении главное, второстепенное и т.д., благодаря чему в его сознании «выстраиваются» ментальные многоуровневые структуры.

Анализ следует начинать с выявления *проблемной ситуации*, породившей коммуникативный акт, и того *социокультурного фона*, на котором он произошёл. Проблемная ситуация «пронизывает» все уровни структуры; это те «строительные леса», которые делают возможным «возведение здания», с одной стороны, порождение коммуникативного акта, с другой — возможность его адекватного истолкования. Интенциональная (мотивационно-целевая) структура всегда выделяется как некое организующее начало, как некий вектор, обнаружи-

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

ваемый при соответствующих навыках анализа (неинтенционального социального общения попросту не бывает). Специалисты, владеющие такими навыками, из одного и того же произведения выделяют одинаковые структуры; как известно, обоснованностью метода является возможность повторяемости опыта.

Как правило, интенция, определяющая и организующая содержательный материал, не выражена словесно: даже *цели общения* не всегда декларируются, в силу разных причин, а *мотивы общения* — тем более, поскольку последние относятся к эмоционально-экспрессивно-образной сфере личности и далеко не всегда оказываются достоянием сознания. «Непроявленные» мотивы, например, типичны для художественного творчества; скрытая, «завуалированная» мотивация — при желании воздействовать, манипулировать.

Мыслительные процессы человека не аналогичны логическому и временному разворачиванию коммуникативного акта — в сознании воспринимающей личности происходят неоднократные перемещения по его виртуальной структуре в удачных или менее удачных попытках «считать», понять мотивы и цели автора. Представляется возможным провести параллель между механизмами поиска интенциональности в коммуникации с *механизмами понимания* в реальной жизни: и в первом, и во втором случаях мы встречаемся со *знаками*, на которые каким-то образом реагируем, пытаемся постичь. В *механизмах понимания* выделяется *дедукция* (рассуждение от посылки к следствию) и *апдукция* (рассуждение от следствия к посылке), которую американский исследователь Ч. Пирс назвал гипотезой, возникающей как озарение: «Конечно, различные элементы этой гипотезы присутствовали в моем сознании и раньше, но именно мысль сопоставить то, что раньше я не подумал бы сопоставлять, заставляет новое предположение вдруг молнией вспыхнуть в моем сознании»¹. Признавая апдукцию как «ненадежный метод — метод выдвижения гипотез, которые опровергаются так же легко, как и выдвигаются», Л.Г. Ионин считает, что иного пути, тем не менее, нет: «дедукция не дает нового знания, в отличие от апдукции. В процессе взаимодействия, природа и тип которого *уже ясны*, могут фигурировать и дедукция, и индукция. Но *опознание* типа нового взаимодействия, сопоставление нового, эмпирического факта с имеющимся набором типов ситуаций, личностей, мотивов, зафиксированным в опыте культуры, в языке, происходит путем апдуктивного заключения»².

Метод интенционального (мотивационно-целевого) анализа основан как на дедуктивных, так и, преимущественно, на апдуктивных механизмах познания, дополняющих и уточняющих умозаключения исследователей. Полученные данные о реальной интенциональности социально значимых коммуникативных актов представляют интерес для представителей самых разных дисциплин и видов деятельности: социологов, психологов, журналистов, аналитиков, критиков, искусствоведов, литературоведов, переводчиков, авторов, читателей и т.д. Например, можно сделать доказательные выводы о таких характеристиках ставших предметом острых общественных дискуссий произведений, как оригинальность авторских мотивов и целей, их соответствие общепринятым морально-нравственным нормам, степень гуманитарной направленности, наличие и особенности манипулятивных приемов.

Социальные технологии с использованием семиосоциопсихологических методов и подходов предусматривают, как правило, не только мероприятия по выявлению данных об интенциональности изучаемых коммуникативных актов (произведений, материалов и т.д.), но и комплекс взаимозависимых процедур, дающих возможность прослеживания (диагностики) инициированных ими социальных последствий. Для этих целей проводится *комплексный междисциплинарный опрос*, позволяющий узнать не только о социально-демографических характеристиках респондентов и их интересах и предпочтениях, но и об особенностях понимания ими смысловых доминант изучаемых коммуникативных актов, а также об их эмоциональных и поведенческих реакциях в связи с восприятием (телепередачи, статьи, романа, выступления и т.д.)³.

¹ Pierce Ch. Collected papers. Cambridge (Mass.), 1956. Vol. 6. P. 181.

² Ионин Л.Г. Социология культуры. М.: Логос. 2000. С. 97.

³ Адамьянц Т.З. Социальные коммуникации. М.: Дрофа, 2009.

Используется многоуровневая анкета, включающая несколько групп вопросов: во-первых, традиционную паспортную часть, выявляющую социально-демографические данные респондентов; во-вторых, группу открытых вопросов с просьбами об интерпретировании тех или иных, согласно задачам исследования, произведений (анализ этой группы вопросов позволяет выявить степень понимания авторской интенциональности); в-третьих, социологические и социально-психологические вопросы, тесты, шкалы и т.д., позволяющие получить данные о значимых характеристиках когнитивной, эмоциональной, профессионально-деловой и других сфер личности (компоненты *социальных представлений*, или «картин мира»).

Безусловную ценность для социального исследователя представляет возможность проследить особенности «преломления» в сознании личности мотивационно-целевой структуры воспринятого коммуникативного акта, степень понимания интенциональности. Для этих целей, как уже упоминалось, в анкету включается «набор» просьб и заданий проинтерпретировать воспринятый материал. Поскольку комплекс полученных в анкете ответов представляет собой целостный, завершённый акт интерпретирования (изучаемого коммуникативного акта), возникает возможность составить мотивационно-целевую структуру восприятия. Сопоставление данных о мотивационно-целевой структуре, исходящей от коммуникатора, со структурой этого же коммуникативного акта, отраженной в сознании респондента, позволяет судить о степени адекватности понимания им интенциональности коммуникатора: насколько адекватно интерпретирует, различает ли главное, второстепенное, третьестепенное и т.д. (по отношению к интенциональности), запомнил ли основные логические и эмоциональные «узлы». Следует особо подчеркнуть, что речь в данном случае идет не о согласии или несогласии с авторской позицией — только о понимании.

При получении достаточного количества сопоставимых результатов (обычно анализируется восприятие не одного, а нескольких произведений) можно дифференцировать опрошенный массив по социоментальным группам (этому понятию синонимичны термины «группы сознания», «интерпретационные группы», «группы по уровню развития коммуникативных навыков» и т.д.). Все эти термины используются для обозначения особенностей ориентации человека в коммуникационных процессах, степени адекватности понимания им интенциональности (основных смысловых доминант) коммуникатора.

В соответствии с принятой классификацией, особенности понимания и интерпретирования авторской интенциональности дифференцируются как *адекватные* (высокий уровень коммуникативных навыков), *частично адекватные* (средний уровень коммуникативных навыков) и *неадекватные* (низкий уровень коммуникативных навыков), а наиболее оптимальным считается такой вариант взаимоотношений между участниками коммуникации, когда обе стороны ориентированы на взаимопонимание и реально его достигают.

Возможность выделения социоментальных групп позволяет «развести» такие понятия, как «смысл коммуникативного акта» и «представления о действительности» в связи с его восприятием. Смысл конкретного коммуникативного акта, или его интенциональность, всегда имеет константный характер, в то время как представления человека о действительности в связи с восприятием, конечно же, имеют (и не могут не иметь) личностные нюансы и оттенки (смыслы). Платформа для конструктивного взаимодействия, тем не менее, возникает только при взаимопонимании между общающимися сторонами, то есть при адекватном восприятии того *самого главного*, что хотят сказать, передать, выразить участники коммуникации.

Возможность выделения социоментальных групп позволяет также углубленное понимание весьма распространенных вариантов восприятия, ставших предметом наблюдения в концепциях *символического интеракционизма* (в рамках этой теории декларируется идея «порабления» людей, живущих в выдуманном ими мире проблем, посредством выдуманных ими же знаков и символов), и *казуальной атрибуции* (здесь декларируется множественность и личностный характер смыслов воспринятых произведений и, соответственно, непознаваемость истины).

Такие особенности восприятия действительно свойственны очень большим группам аудитории: это группы *частично адекватного* и *неадекватного восприятия*. Не стоит, однако,

списывать со счетов и более редкий (и наиболее ценный) вариант, при котором *адекватно* понятая интенциональность коммуникатора приобретает в сознании воспринимающей личности статус реального, объективного факта (о том, каковы на самом деле цели и мотивы автора) и только затем вписывается в «картину мира», осознанно влияет (или не влияет) на характеристики представлений о действительности, с учетом жизненного опыта, знаний, намерений и т.д.

Процедура социально-диагностических работ предусматривает также сопоставление полученных данных об особенностях понимания и интерпретирования респондентов (принадлежности к той или иной социоментальной группе) с традиционными социологическими и социально-психологическими сведениями, также получаемых в результате опроса, благодаря многоуровневой анкете. Можно сделать выводы, например, об особенностях эмоциональных и поведенческих реакций в связи с воспринятыми фактами, событиями, проблемами, персоналиями не только социально-демографических и социально-психологических групп, как это происходит при традиционных вариантах изучения, но и групп социоментальных, что, как правило, позволяет обнаружить тенденции и закономерности, не фиксируемые при других вариантах анализа.

Тенденция к определяющему влиянию уровня развития коммуникативных навыков на качество (социальную значимость) представлений современной молодежи о действительности была зафиксирована и рамках проекта РФФИ «Развитие коммуникативных навыков личности в зависимости от степени диалогичности социокультурной среды» (2008–2010 гг.). Как известно, сходные интересы, предпочтения, реакции и т.д. можно встретить в разных, порой прямо противоположных по социально-демографическим и социально-психологическим характеристикам молодежных группах, и разнонаправленные — в социально-гомогенных группах, однако найти индикаторы, определяющие такую ситуацию, удалось только в результате использования комплексных социально-диагностических технологий.

Представления о действительности группы с высоким уровнем коммуникативных навыков (адекватное восприятие) современной молодежи, по свидетельству исследовательских данных, выгодно отличаются от представлений о действительности других групп: характеристики личности знающей, творческой, ориентированной на взаимопонимание и конструктивные взаимодействия, ответственной в социальной и профессиональной сферах неразрывно связаны с высоким уровнем коммуникативных навыков.

В соответствии с исследовательскими данными, представители этой группы чаще, нежели представители других социоментальных групп, «точкой отсчета» при самоидентификации (при просьбе рассказать о себе) называют свою принадлежность стране, планете Земля; их личностные представления о желательном (что важно, чего хотелось бы...) чаще, нежели в других группах, связаны с общечеловеческими ценностями; их личностные представления о нежелательном (чего не хотелось бы...) гораздо чаще, чем в других группах, связаны с нежеланием помех для общечеловеческих ценностей; в эмоциональном восприятии «мира вокруг» здесь преобладают нейтральные тона, а пессимистичных тонов меньше, чем в других группах; характеристики любимых персонажей здесь преимущественно социально позитивные (добрый, смелый, мужественный, патриот и т.д.); большая часть представителей этой группы обнаружили в своих многочисленных ответах высокий уровень социокультурных интересов (они чаще, чем представители других социоментальных групп, читают газеты и журналы, смотрят телевизор, обращаются к Интернету, помнят большее количество названий кинофильмов, фамилий телезвезд; среди телеведущих, которых они называют любимыми, представители разных жанров, в том числе информационных, аналитических; большая часть (работающих в крупных корпорациях) представителей этой группы адекватно интерпретируют внутрикорпоративные нормы, законы и правила, неадекватных вариантов интерпретирования внутрикорпоративных норм, законов и правил здесь вообще нет, в отличие от других групп; большинство представителей этой группы основной приоритетностью среди внутрикорпоративных норм, законов и правил называют престиж страны, глобальные интересы корпорации; эмоциональное отношение большей части представителей этой группы к

внутрикорпоративным нормам, законам и правилам позитивное¹. Полученные результаты не только убедительно свидетельствуют о преимуществах представителей группы с высоким уровнем коммуникативных навыков, но и актуализируют задачу массового развития качества сознания людей, наших современников, прежде всего детей и молодежи, что обеспечит гармоничное развитие общества и каждой отдельной личности.

Социально-диагностический анализ может также включать работу по систематизации выявленных в процессе исследовательской работы мнений, точек зрения, предложений, реакций и т.д. в зависимости от уровня развития коммуникативных навыков респондентов и приданию им разного «статуса». Логика подсказывает, например, что при принятии управленческого решения оценки, мнения и предложения группы, имеющей высокий уровень развития коммуникативных навыков (адекватное восприятие), должны иметь больший вес, нежели оценки, мнения и предложения других социоментальных групп. Важно, однако, учитывать реакции и групп со средним и низким уровнями развития коммуникативных навыков (частично адекватного и неадекватного восприятия), поскольку они составляют большинство населения (аудитории, электората и т.д.).

Еще один возможный аспект социально-диагностических работ — прослеживание социальных последствий, инициированных тем или иным произведением (материалом) и в целом доминантными характеристиками социокультурной среды. Так, при изучении социокультурной среды современных детей и молодежи неоднократно фиксировалось, что мотивационно-целевые доминанты ряда широко рекламируемых и потому популярных произведений «перехлестывают» за те морально-нравственные рамки, которые традиционно узаконены обществом, наукой, педагогикой, классическим искусством, религией. Это и поэтизация brutальных отношений, агрессивности и жестокости, неразборчивости в средствах достижения цели, волшебных способностей и тайных знаний, дающих могущество и власть. «Безболезненно» встречи с такими произведениями проходят только при развитых навыках понимания, а это качество, напомним, присуще только небольшой части аудитории; представители же остальных социоментальных групп далеко не всегда отличают вымысел и авторские фантазии от реальности и стремятся походить на неоднозначных по своим морально-нравственным качествам, но «крутых», «прикольных» (лексика взята из анкет) персонажей².

В общественном мнении о подобных произведениях существует разные точки зрения: с одной стороны, имеется значительное число их противников (здесь педагоги, воспитатели, родители и т.д.), с другой — защитники с утверждениями о модности, современности, востребованности такого «формата» и ретроградности противников его широкого распространения. Арбитром в таком противостоянии мнений должна стать в первую очередь социальная наука, дающая обоснованные ответы о реальном социальном последствии подобных произведений для разных групп аудитории, прежде всего для детей и молодежи, а также и конкретные рекомендации по проектированию социокультурной среды для тех, кто в них нуждается (следует особо подчеркнуть, что речь не идет о цензуре, но о понимании тенденций и последствий коммуникационных процессов). Такие данные необходимы и для общественных и экспертных советов, и для представителей управленческих структур, и для педагогов, воспитателей, родителей.

В зависимости от поставленных задач, в социально-диагностических и социально-проектных исследованиях применяется тот или иной «набор» взаимосвязанных исследовательских стратегий и процедур. Так, при реализации вышеупомянутого проекта РФФИ были проведено несколько этапов изучения, в следующей последовательности и сопряженности:

1. Составление интенциональных (мотивационно-целевых) структур произведений, ставших предметом анализа. Эта процедура позволила выявить подлинную интенциональность

¹ Адамьянц Т.З. Уровень развития коммуникативных навыков как индикатор в процессах самоопределения и самоорганизации личности // Мир психологии. 2010. № 3. С. 253–261.

² Адамьянц Т.З. Дети в современной социокультурной среде // Россия реформирующаяся. Ежегодник / Отв. ред. М.К. Горшков. Вып. 8. М.: ИС РАН, 2009. С. 407–421.

произведений, сделать выводы об ее социальной значимости, соответствии морально-нравственным нормам, степени оригинальности и способах донесения.

2. Проведение комплексного опроса аудитории, позволяющего получить, во-первых, данные о социально-демографических характеристиках респондентов; во-вторых, данные об особенностях понимания и интерпретирования изучаемых произведений, что позволило дифференцировать опрошенный массив респондентов по уровню развития коммуникативных навыков; в-третьих, данные о характеристиках их представлений о действительности («картин мира»).

3. Поиск взаимозависимостей между уровнем развития коммуникативных навыков опрошенных респондентов и характеристиками их представлений о действительности («картин мира»).

4. Фиксирование коммуникативных сбоев, выражающихся в непонимании респондентом авторской интенциональности.

5. Поиск причин коммуникативных сбоев – и на уровне структурной организации материала, и на уровне особенностей восприятия личности.

6. Социальная диагностика социокультурной среды, а также отдельных популярных книг, газет, журналов, теле- и радиопрограмм, интернет-сайтов, мультфильмов, кинофильмов, компьютерных игр, произведений эстрады, театра (выявление подлинной интенциональности, поиск тенденций в особенностях влияния на «картины мира» респондентов, с дифференциацией по уровню развития коммуникативных навыков, прогнозирование возможных социальных последствий при восприятии разными группами аудитории).

7. Выдача рекомендаций по созданию коммуникационных потоков, способствующих организации здоровой, социально ориентированной социокультурной среды.

Начало комплексных междисциплинарных социально-диагностических исследований было положено в широко известном исследовательском проекте «Общественное мнение», который был реализован в 1969–74 гг. под руководством Б.А. Грушина в г. Таганроге (опрошено 300 человек). Именно тогда в социальной науке был заявлен термин «группы сознания» и обозначены параметры этих групп (при восприятии печатных материалов СМК общественно-политического содержания). Устойчивое, адекватное понимание печатных материалов СМК общественно-политического содержания было зафиксировано всего у 14% участников эксперимента¹. При этом в их числе были и те, кто обнаружил высокий уровень знания соответствующей лексики, персоналий и т.д. (12 %), и, частично, те, у кого аналогичные показатели тезауруса оказались низкими (2 %).

Несмотря на бесспорную социальную значимость подобного знания, изучение особенностей восприятия (понимания и интерпретирования) общественно-политических и информационных материалов СМК проводится крайне редко. В 1980-ые годы изучение восприятия материалов газеты «Правда», с использованием тех же методик и тех же текстов для восприятия, что и в проекте «Общественное мнение», провел А.В. Жаворонков. Результат оказался таким же — 14 % адекватных интерпретаций². В 1990-е годы, при изучении особенностей интерпретирования телепередач общественно-политического содержания, устойчивое адекватное восприятие показали 13,6 % телезрителей³.

Такая константность в показателях, несмотря на значительные интервалы между опросами и произошедшее за этот период расширение возможностей для аудитории получения информации, говорит о стабильности и устойчивости социоментальных характеристик людей.

В рамках проведенных исследований, кроме сведений о параметрах социоментальных групп, были получены и другие ценные исследовательские данные: особенности понимания общей и специальной лексики, штампов, устойчивых словосочетаний, знание персоналий и т.д.

¹ Массовая информация в советском промышленном городе. Опыт комплексного социологического исследования / Под ред. Б.А. Грушина и Л.А. О니кова. М.: Политиздат, 1980.

² Жаворонков А.В. Российское общество: потребление, коммуникации и принятие решений, 1967–2004 гг. М.: Вершина, 2007.

³ Адамьянц Т.З. К диалогической телекоммуникации: от воздействия — к взаимодействию. М.: ИС РАН, 1998.

При изучении особенностей восприятия общественно-политических телепрограмм (1990-е годы) была прослежена взаимозависимость между уровнем развития коммуникативных навыков опрошенных респондентов и их представлениями о действительности в сфере политической и социальной жизни. При этом только немногочисленная группа аудитории, обладающая высоким уровнем развития коммуникативных навыков (*адекватное восприятие*), оказалась не подверженной манипулятивным приемам. Интенциональность телеведущего или участника телепрограммы, которую они «видят», то есть адекватно понимают, становится для них фактом объективной реальности, знание о котором входит составной частью в их «картины мира». Выводы и решения относительно воспринятых материалов или проблем эти люди принимают самостоятельно, при этом оставаясь (в процессе опроса) выдержанными и корректными.

Представители группы *частично адекватного восприятия* (не менее 30–35%, в зависимости от формы подачи и организации материала) хорошо запоминали содержание воспринятых телепередач. Эти люди, как правило, были хорошо информированы по многим вопросам, но мотивы и цели коммуникатора все же оставались для них в «смысловом вакууме», они их попросту не замечали. Для желающих манипулировать успешной в этой «группе сознания» оказывается интенция «все плохо», вызывающая повышенную тревожность, страхи и, соответственно, желание следовать рекомендациям о том, что или как следует сделать (поступить, отреагировать).

Представители группы *неадекватного восприятия* (также порядка 30–35% аудитории в зависимости от формы подачи и организации материала) отличались либо взрывной эмоциональной реакцией по поводу «выхваченных» из контекста сведений и фактов, либо отсутствием интереса, скукой, вялостью. При этом и в первом, и во втором случаях фиксировалось полное непонимание основных смысловых доминант воспринятых материалов, нежелание или неумение «включить» внимание и попытаться понять. Эта группа также поддается манипуляции; здесь эффективны скрытые психологические приемы, побуждающие к эмоциональным реакциям и без того эмоционально неустойчивую часть аудитории.

В рамках ФЦП «Формирование установок толерантного сознания личности и профилактика экстремизма в российском обществе, 2001–2005 гг.» были реализованы три самостоятельных проекта с использованием социально-диагностических технологий. Впервые были определены параметры социоментальных групп среди школьников Москвы и Подмосковья, а также зафиксирована взаимозависимость между уровнем развития их коммуникативных навыков, с одной стороны, и, с другой — содержанием и эмоциональной окрашенностью их «картин мира», в том числе степенью толерантности по отношению к сверстникам *другого* вероисповедания, *другой* национальности, и т.д.¹

Изучению социоментальных характеристик аудитории, а также качественных характеристик современной социокультурной среды был посвящен уже упоминавшийся проект РФФИ «Развитие коммуникативных навыков личности в зависимости от степени диалогичности социокультурной среды» (2008–2010 гг., всего опрошено 800 человек). В рамках проекта впервые были определены параметры социоментальных групп среди аудитории всех возрастных категорий, при восприятии художественных произведений: высокий 25 %, средний 47 %, низкий 28 %. Зафиксирована тенденция к повышению уровня развития коммуникативных навыков в связи с возрастом: показатели адекватного восприятия в группе 5–6 лет составляют 12 %, в группе 7–10 лет — 15%, в группе 11–14 лет — 20 %, в группе 15–17 лет — 23 %, в группе 18–25 лет — 25 %, в группе 26–35 лет — 29 %, в группах 36–40 и 41–59 лет — 30 %, в группе старше 60 лет — 38 %.

Полученные результаты подтвердили выводы проведенных ранее исследований о том, что гарантией высокого уровня коммуникативных навыков не являются ни возраст, ни пол, ни место жительства, ни даже уровень образования и род занятий: это особый вид тренированности, навыки, которые без соответствующего природного дара развиваются только при определен-

¹ Дети и проблемы толерантности: Сборник научно-методических материалов / Отв. ред. Т.З. Адамьянц. М.: ИС РАН, 2003.

ных условиях. Проведенная серия экспериментов показала, что на развитие коммуникативных навыков личности плодотворное влияние оказывает взаимодействующая с ней диалогическая коммуникативная среда (субъект-субъектные отношения), а также знакомство с общими принципами мотивационно-целевой организации коммуникативных актов.

В настоящее время комплексные социальные технологии с использованием метода интенционального (мотивационно-целевого) анализа используются в основном либо в рамках некоммерческих инициативных проектов, либо при сборе эмпирического материала для диссертационных исследований¹. И в первом, и во втором варианте полученные результаты оказываются известными преимущественно в узкой научной среде, так же, как и сами технологии. Ни в одной редакции СМК пока нет службы, где проводилось бы изучение социальных последствий выходящих материалов. Практически не анализируется и степень диалогичности используемых творческих приемов, форм подачи материала, личностных мотивационно-целевых доминант, продуцируемых популярными журналистами, телеведущими, эстрадными звездами, артистами, общественными деятелями. Представляется, что такого рода изучение следует проводить постоянно, прослеживая тенденции восприятия различных материалов в конкретном социальном контексте, учитывая в информационной политике не только интересы и ожидания разных групп аудитории, но и возможные социальные последствия. Желателен и социальный заказ на проведение такого рода исследований.

Метафорические интерпретации реформы МПС: восприятие в СМИ и возможности использования технологий речевого воздействия

Баранов Анатолий Николаевич,
Институт русского языка РАН

1. Постановка задачи

Одна из важнейших задач прикладной политологии и политической лингвистики заключается в исследовании восприятия феноменов политической жизни в общественном сознании. Рассмотрим эту задачу на примере реформы МПС, планировавшейся и осуществлявшейся в начале 2000-х годов. Эта реформа неоднозначно воспринималась в российском обществе. В данной работе проводится анализ восприятия реформы в указанный период времени, и предлагаются направления корректировки образа этой реформы с помощью техник речевого воздействия с использованием метафор, которые могли бы способствовать улучшению этого образа.

В когнитивной теории метафоры утверждается, что метафора оказывает существенное влияние на процесс принятия решений человеком. Как известно, процесс принятия решений включает следующие основные этапы: (1) оценка проблемной ситуации; (2) выявление и формирование множества альтернатив решения проблемной ситуации; (3) оценка альтернатив с точки зрения имеющихся ресурсов, ценностных предпочтений и т.д.; (4) выбор альтернативы — принятие решения в узком смысле.

Метафорическое осмысление ситуации оказывается наиболее важным при формировании множества альтернатив решения проблемы. Метафора фокусирует внимание человека на отдельных аспектах проблемной ситуации, коммуникативно высвечивает их и именно в этом смысле влияет на процесс принятия решений. Коммуникативное выделение релевантных свойств ситуации с помощью метафоры называется в когнитивной теории метафоры **профилированием**.

В центре внимания исследований по метафорике дискурса лежит понятие метафорической модели. Под **метафорической моделью** (М-моделью) понимается понятийная область, которая используется для метафорического осмысления некоторой сущности — реалии дейст-

¹ См., например, Гармажапова Е.А. Проблемы эффективности коммуникативных взаимодействий во внутрикорпоративном управлении // Вестник Бурятского государственного университета; серия Философия, социология, политология, культурология. 2009. № 6; Ермизина Е.В. Социальная ответственность СМК в аспекте семиосоциопсихологической парадигмы // Вестник РУДН, серия Социология. 2009. №1.

вительности, феномена общественного сознания и публичной политики. Например, осмысление чего-либо в терминах военных действий, армии, происходит в рамках М-модели войны (ср. *война законов, обоймы аргументов, солдаты политического фронта* и т.п.). На аналогичных основаниях выделяются М-модели родственных отношений, персонификации, механизма, организма, пути-дороги, погоды, фауны, растения-дерева и др.

Метафорические модели отличаются друг от друга метафорическими следствиями, то есть теми свойствами обсуждаемой ситуации (или сущности), которые они профилируют в процессе метафоризации. Так, М-модели войны, игры, спорта профилируют идеи конфликта, победы vs. поражения, ср. *идущая сейчас атака на новые власти, попридержаться рвущихся из окопов коммунистических консерваторов, не садиться за стол переговоров по причине вероятного «проигрыша», запрещение митингов — это капитальный проигрыш Горбачева; играть в международных делах в поддавки*. А М-модели растения-дерева, организма, медицины в политическом дискурсе часто профилируют такую совокупность характеристик, как болезнь, лечение болезни, боль, причиняемая организму (в частности, болезненность оперативного вмешательства). Например, *чума роста цен, политическое слабоумие, метастазы большевизма, эпидемия суверенизации, шоковая терапия*.

Функционирование языка в обществе происходит не одинаково в различных проблемных областях и социальных стратах. Хотя литературный язык и задает некоторые общие рамки использования русского языка, тем не менее, каждая профессиональная сфера и конкретная совокупность носителей привносит свои специфические черты во множество речевых высказываний. Различающиеся по тематике и социальному составу носителей области использования языка называются **дискурсами**. Здесь можно говорить о «тоталитарном дискурсе», «дискурсе безопасности», «дискурсе свободы и справедливости», «парламентском дискурсе». Именно в этом смысле можно говорить о «дискурсе реформы МПС»: это совокупность текстов СМИ, в которых обсуждается проблематика реформы МПС и частично проблематика реформирования монополий в целом.

Совокупность М-моделей, использующихся в данном дискурсе, образуют **метафорику** данного дискурса. Часть из этих М-моделей типична для данного дискурса, формируют его как таковой, как специфический и отличающийся от других типов дискурсов. Таким образом, исследование метафорики дискурса о реформе МПС предполагает выявление основных М-моделей, в терминах которых осмысливается реформа МПС. Кроме того, необходимо выявить семантические связи между отдельными М-моделями. Это даст возможность определить, устойчив ли данный дискурс с когнитивной точки зрения. Наконец, важно установить типичные свойства, которые профилируют частотные М-модели в обсуждаемом типе дискурса.

2. Эксперимент и его основные этапы

Проведенный анализ основывался на корпусе текстов СМИ по проблеме реформы МПС, объемом 3.400.000 словоупотреблений¹. Основная тема документов корпуса (далее Корпус) — реформа МПС и других естественных монополий. Три с половиной миллиона употреблений представляют собой достаточно репрезентативный массив, поскольку классический Брауновский корпус американского английского включает 1 миллион словоупотреблений. Тексты Корпуса были обработаны программой «КСИ» по технологии контент-анализа. В качестве концептуальной переменной выступило слово *реформа* (в различных падежных формах). Полученные контексты были введены в базу данных и проанализированы с точки зрения использования метафор, точнее, метафорических моделей.

Всего из порядка 15 тысяч контекстов употребления лексем *реформа* был выделен 681 контекст использования метафор по отношению к понятию «реформа». Каждый из контекстов был кодирован с помощью метаязыка семантических дескрипторов². Далее с помощью

¹ Корпус текстов формировался на основе данных, предоставленных информационно-аналитической службой Медиалогия.

² О концептуальном инструментарии дескрипторной теории метафоры см., в частности: Баранов А.Н. Метафорические модели как дискурсивные практики // Известия РАН. Сер. литературы и языка. 2004. Т. 63. № 1. С. 33–43.

программы HyperNav, разработанной в отделе экспериментальной лексикографии Института русского языка РАН, проводился подсчет метафорических моделей с помощью дескрипторов разных уровней обобщения. Обратимся к полученным результатам.

3. Метафорические модели дискурса о реформе МПС

В Таблице 1 представлены метафорические модели дискурса о реформе МПС, упорядоченные по убыванию частоты употребления. (Опущены М-модели, абсолютная частота употребления которых меньше 8.) Опишем базовые М-модели, с входящими в них М-моделями более низкого уровня абстракции, а также выявим типичные свойства реформы, которые профилируются этими М-моделями.

Таблица 1

Метафорические модели дискурса о реформе МПС (различного уровня абстракции)

Метафорическая модель	Абсолютная частота	Относительная частота
война	195,00	0,117047
армия	169,00	0,101441
штаб	169,00	0,101441
объект-предмет	154,00	0,092437
движение	154,00	0,092437
движущийся объект	132,00	0,079232
персонификация	120,00	0,072029
транспорт	76,00	0,045618
транспортное средство	74,00	0,044418
чел-в пути	49,00	0,029412
путь-дорога	32,00	0,019208
пространство	29,00	0,017407
механизм	20,00	0,012005
ограниченное пространство	17,00	0,010204
спорт	15,00	0,009004
военные действия	14,00	0,008403
автоимное употребление	14,00	0,008403
рамки	13,00	0,007803
шаг	13,00	0,007803
мотор	12,00	0,007203
строение	12,00	0,007203
спортсмен	11,00	0,006603
растение-дерево	11,00	0,006603
плод	9,00	0,005402
поезд	8,00	0,004802

Начнем с наиболее частотной М-модели «война». Всего в выборке фиксировано 195 употреблений этой модели, причем на «армию» и «штаб», как более частные метафоры, входящие в М-модель войны, приходится 169 употреблений. Метафора армии представлена только метафорой штаба (*штаб по проведению реформы*). Это объясняется тем, что метафора армии восстанавливалась как промежуточный дескриптор в базе данных для логичности перехода к М-модели войны. Кроме указанных более частных метафор, в М-модель войны входит также метафора военных действий (*реформы естественных монополий не делают **наскоком***; *командный состав МПС «сверил часы» перед началом реформы железнодорожной отрасли* и т.п.). Типичные свойства реформы, которые профилируются М-моделью войны — это оперативность, эффективность, быстрота проведения реформы решимость в ее осуществлении (*реформы естественных монополий не делают **наскоком***; *До конца ноября правительство обсудит базовый пакет законопроектов, на основе которых должна проводиться реформа МПС. «Реформаторский штаб» министерства их уже подготовил и даже обсудил с участниками рынка за двумя «круглыми столами» — в Российском союзе промышленников и предпринимателей и Торгово-промышленной палате*). При этом М-модель

войны предполагает такие рамки обсуждения и концептуализации реформы, в которых реформа воспринимается как конфликт, как сражение, которое нужно выиграть: *В скорую реформу отечественных стальных магистралей мало кто верил. Ведь первоначальный проект реформ попал под прицел критики. Обстрел велся так плотно, что казался заградительным.* [Экономика и Время (Санкт-Петербург), 13. 01. 2003]

Вторая метафорическая модель, характерная для дискурса о реформе МПС, — модель объекта-предмета (154 вхождения). Она включает в себя более частную метафору движущегося объекта (132 контекста). Основное свойство, профилируемое данной М-моделью, — концептуализация реформы, как предмета, который можно задвинуть (*реформа вообще будет "задвинута"*), отодвинуть (*Рассуждать же о том, нужна ли реформа сегодня или же ее отодвинуть на завтра, - значит тянуть кота за хвост.*), отложить (*реформа, как и три года назад, будет надолго отложена*). Другой аспект опредмечивания реформы связан с профилированием ее как объекта, который движется (*Иначе реформа не двинется дальше стадии предварительных проработок; Реформа МПС медленно движется к своему концу*).

Интерпретация реформы МПС как движущегося объекта естественно сочетается с М-моделью движения (154 вхождения) и с М-моделью транспорта (76 вхождений). Метафора движения высвечивает свойство реформы МПС как процесса, предавая это семантикой движения: *движется/не движется структурная реформа отрасли*.

М-модель транспорта имеет две более частные метафоры: «транспортное средство» и «поезд». Высокая частота употребления этой М-модели объясняется двумя факторами. Во-первых, метафора транспортного средства семантически (по метафорическим следствиям) тесно связана с М-моделью движения и объекта-предмета (через метафору движущегося объекта). Во-вторых, сам предмет обсуждения, реформа железнодорожной отрасли, стимулирует использование по сути игровых употреблений, когда реформа МПС интерпретируется как поезд. Ср. *События последнего времени свидетельствуют: скорый поезд реформы МПС набирает ход; Поезд реформы тронулся; Поезд реформ на железнодорожном транспорте наконец-то вышел из депо; Чтобы поезд реформ не терял вагоны*. Использование метафоры поезда в основном профилирует идею начала реформы и процесс ее реализации.

Одна из типичных возможностей профилирования свойств реформы МПС в рамках метафоры транспортного средства касается выбора направления движения и реализуется в М-модели пути-дороги (32 вхождения). Ср. *зажегся «зеленый» на пути структурной реформы железнодорожного транспорта*. Метафора пути, по которому движется реформа, или пути как реформы профилирует также идею различий в подходе к реформе (*существуют разные мнения о направлении структурной реформы; какие пути реформ видятся профессионалам*), идею реформы как процесса (*решить еще одну важную проблему на пути реформ; перевод железнодорожного транспорта на рельсы реформ*), возможные проблемы в процессе реализации реформы (*с тревогой вглядываются вперед локомотивщики: как бы по пути реформ не проскочить на красный*); сам факт проведения реформы (*отрасль действительно встала на путь реформ*).

Из наиболее частотных М-моделей следует упомянуть также метафору персонификации (олицетворения), которая позволяет категоризовать неживые сущности как живые и одушевленные. В рассмотренной выборке М-модель персонификации позволяет подчеркнуть действующий характер реформы (*реформа напрямую повлияет на судьбы полутора миллиона железнодорожников*); идею необходимости проведения реформы, ее императивность (*реформа должна кардинально изменить ситуацию*); целевую составляющую реформы (в конечном счете к увеличению зарплаты, к чему и *стремится реформа; реформа обеспечит железным дорогам РФ дополнительные инвестиции; реформа МПС не урежет социальные льготы железнодорожников*); возникновение проблем в процессе реформирования (*Сколько вопросов поставила реформа!*).

М-модель спорта (15 употреблений; метафора спортсмена 11 употреблений) в политическом дискурсе чаще всего профилирует идею конфликта через семантику соревнования, од-

нако в собранном массиве данная модель связана с метафорой персонификации и М-моделью движения, при этом профилируется идея начала/конца реформы: *реформа железнодорожной отрасли на старте; реформа Министерства путей сообщения вышла на финишную прямую; реформа МПС уже где-то у старта.*

М-модель растение-дерево представлена 11 употреблениями, 9 из которых приходится на метафору плода. Основное свойство, профилируемое данной моделью — это идея необходимости реформы МПС: *назрела реформа отрасли; реформа в системе МПС назрела.*

Все обсужденные М-модели характеризуются одним общим свойством: при осмыслении реформы МПС по этим метафорическим моделям реформа МПС предстает как непрозрачная сущность. Иными словами, такие способы метафоризации оставляют за рамками обсуждения содержание реформирования. Реформа интерпретируется как что-то, что начинается, завершается, протекает с большей или меньшей скоростью, имеет какие-то цели и задачи, имеет проблемы, варианты воплощения. Но сама реформа остается в дискурсе «черным ящиком», не доступным взгляду простого смертного.

Тем не менее, в дискурсе о реформе МПС обнаруживаются и рациональные М-модели, которые в принципе позволяют осмыслять ее с более конструктивных позиций — как относительно прозрачную сущность, состоящую из нескольких компонентов. К таким метафорическим моделям относятся М-модели механизма (20 употреблений, в том числе с метафорами мотора и строения по 12 употреблений). Однако проанализированный материал показывает, что даже при использовании этой рациональной по сути М-модели профилируются смыслы, связанные с функционированием механизма, а не с его компонентным составом. Иными словами, коммуникативно высвечиваются идеи начала проведения реформы и процесса реформирования (*Администрация края, создав свою компанию-оператора, заявила, что начала действовать, не дожидаясь того момента, когда реформа закрутится и поправить что-то будет сложно*). Отсюда следует, что и рациональная М-модель механизма, используется в дискурсе о реформе МПС не для осмысления рационального поведения. Содержание реформы МПС как таковой остается непрозрачным.

М-модель строения также относится к числу рациональных метафор. И в отношении этой М-модели ожидания рациональной интерпретации реформы МПС оправдываются. Ср. примеры типа *пакет законопроектов, на которых будет базироваться реформа МПС; реформа, краеугольным камнем которой является создание четкого законодательства; правовой каркас структурной реформы отрасли*. Следует, однако, отметить, что и при использовании метафоры строения выбирается довольно ограниченный набор элементов строения: база (базироваться), каркас, краеугольный камень.

Итак, общий анализ употребления М-моделей дискурса о реформе МПС с определенностью указывает на то, что подавляющая часть метафорики используется не для обсуждения содержания реформы как таковой, а для характеристики ее свойств как единого и непрозрачного для адресата целого, феномена, аналогичного «черному ящику», который доступен независимому наблюдателю только по информации на входе ящика и на его выходе. Метафоры высвечивают такие характеристики реформы, как ее начало и конец, возможности реализации, ее процессуальный характер, необходимость реформирования, быстроту осуществления, возможность различных способов проведения реформы, возможность ее полного или частичного прекращения. При этом составляющие реформы, ее содержание не поддерживаются метафорическими интерпретациями.

4. Дискурс о реформе МПС: устойчивость состава метафорических моделей

Не все выделенные частотные М-модели семантически связаны между собой. Здесь можно выделить несколько относительно автономных блоков. Один из важных блоков — М-модель войны, которая в исследованной выборке не обнаружила семантических пересечений с другими метафорами.

Второй блок сформирован М-моделью объекта-предмета (с метафорой движущегося объекта). Метафора движущегося объекта связывает М-модель объекта-предмета с М-моделью движения и транспорта. В свою очередь, метафоры движения и транспорта оказываются свя-

заны с метафорой пути-дороги и, далее, с М-моделью пространства, поскольку любое движение транспортных средств происходит в пространстве. В силу того, что по пути-дороге и в пространстве может двигаться и человек (ср. контексты типа *Куда пойдут реформа МПС?*), эти метафоры семантически связаны с М-моделью персонификации. В свою очередь, персонификация (и метафора движения) связана по метафорическим следствиям с М-моделью спорта (прежде всего через метафору спортсмена). Отдельные блоки образует сравнительно низкочастотные М-модели растения-дерева, строения и механизма.

О чем говорит такая структура метафорического дискурса? Если пренебречь блоками низкочастотных М-моделей (растение-дерево, строение и механизм), то остаются два основных блока: один образован метафорической моделью войны, а второй — комплексом связанных между собой разнообразными метафорическими следствиями М-моделей объекта-предмета (с метафорой движущегося объекта), движения, транспорта, пути-дороги, пространства, персонификации и спорта. М-модель войны представляет собой чрезвычайно конфликтный способ осмысления действительности, а М-модели, входящие во второй блок неконфликтны или слабоконфликтны. Поскольку и М-модель войны и М-модели второго блока высокочастотны (195 и 580), то **налицо противоречие, которое создает когнитивную неустойчивость дискурса о реформе МПС.**

5. Реформа МПС: возможности речевого воздействия

Когнитивная неустойчивость метафорической структуры дискурса о реформе МПС в рассматриваемом случае говорит о том, что имеются альтернативные точки зрения на реформу МПС, которые противоборствуют между собой.

В метафорической структуре видны следы рг-овских мероприятий. Это хорошо определяется на примере использования полустертой метафоры штаба по отношению к структуре, отвечающей в МПС за разработку плана реформирования. Понятны метафорические следствия, которые имели в виду авторы номинации: выделить, усилить и донести до воспринимающей аудитории идею оперативности, четкости и действенности работы данной структуры. Однако одновременно они ввели идею конфликтности, которая изначально практически отсутствовала в осмыслении концепции реформирования МПС. Фактически единственной относительно конфликтной метафорической моделью, кроме М-модели войны, которая представлена в исследованном дискурсе, является М-модель спорта (15 употреблений), входящая во второй блок М-моделей. Очевидно, что ни по каким критериям эта М-модель не может претендовать на статус высокочастотной по отношению к реформе МПС. Широкое использование метафоры штаба привносит конфликтность в дискурс, который в сущности малоконфликтен.

Интересно, что если опустить метафору штаба, оставив только военные метафоры типа *авангард, позиция, прорыв, попасть под обстрел критики, наскок* и т.п., то метафорическая структура дискурса существенно меняется в сторону уменьшения конфликтности, причем М-модель войны теряет положение самой частотной и перемещается с первого места на седьмое. Искусственная конфликтность дискурса по реформе МПС устраняется, что существенно повышает его приемлемость для адресата и, как следствие, успешность в формировании позитивной оценки реформы. Кроме того, устранение количественной доминанции М-модели войны позволит достичь когнитивной устойчивости метафорики, используемой для осмысления и позиционирования реформы.

Отсюда очевидное соображение: **название структуры, отвечающей в МПС за разработку планов по реформированию отрасли, следовало бы изменить.** Оно не должно было восприниматься как элемент конфликтной коммуникации.

В метафорических осмыслениях реформы МПС было представлено всего 11 базовых М-моделей. Как было показано в целом ряде исследований¹, активное использование различных метафорических моделей и возрастание частоты метафор в рамках конкретной М-модели

¹ См., например: Баранов А.Н. Метафоры в политическом дискурсе: языковые маркеры кризисности политической ситуации // *Linguistic Change in Europe 1990–2000* / Ed. by L. Zybatow. Wien, 2000.

указывает на кризисность осмысления соответствующего феномена. В рассматриваемом случае обнаруживается узость спектра использованных метафор, что может свидетельствовать о том, что в общественном сознании реформа МПС **не воспринимается как критическая величина**, как политический или политико-экономический кризис.

Отчасти этот феномен объясняется недостатком информации о проведении самой реформы. Изменения на железнодорожном транспорте, прежде всего, рост цен на билеты, воспринимаются не как часть реформы, а как часть общей инфляционной составляющей народного хозяйства. Ничего толком не зная о реформе, трудно оценивать ее как кризис той или иной силы. С этим связано и фактическое отсутствие в исследовавшейся выборке М-моделей, которые обычно используются для профилирования характеристик внутренней структуры объекта метафоризации: механизма и строения.

На набор метафорических моделей существенное влияние оказывает сама сущность объекта метафорического осмысления. При осмыслении реформы железнодорожной отрасли велик соблазн представить реформу как движущийся объект или транспортное средство типа паровоза/локомотива/поезда. И действительно, М-модель объекта-предмета находится на втором месте по частоте употребления после метафоры войны, причем подавляющая часть приходится именно на движущийся объект. Это одна из важнейших характеристик дискурса о реформе МПС, которая вольно или невольно отвлекает внимание от обсуждения сущностных аспектов реформы МПС. Отсюда еще одно соображение, касающееся организации дискурса о реформе: **при позиционировании сущностных аспектов реформирования следует рекомендовать использование М-моделей механизма и строения.**

Анализ показал, что одна из важнейших неконфликтных метафорических моделей — метафора растение-дерево — находится на периферии системы метафор дискурса о реформе МПС. Между тем, воздействующий потенциал этой метафоры весьма значителен. Использование в речевом воздействии **метафоры растение-дерево способствовало бы более благоприятному восприятию аргументации в пользу проведения реформы.**

6. Вместо заключения

Метафора — важный источник анализа состояния общественного сознания и в то же время эффективный инструмент воздействия на мышление человека. Одна и та же ситуация, один и тот же сюжет может быть назван «Побегом, тремя покушениями на убийство и зверской расправой» и русской народной сказкой «Колобок». В первом случае использовалась криминальная метафора, во втором — стандартная номинация известного текста. Анализ метафорики дискурса позволяет не только выявлять отношение общества к тем или иным аспектам политической жизни, но и формировать речевые стратегии влияния на дискурс и, соответственно, мышление. Впрочем, последнее требует сугубой осторожности.

Реконструкция по речи социального статуса говорящего: эвристический потенциал и методологические трудности

Бумагин Роман Евгеньевич,
ЗАО «Рефакта»

Присутствие в концептуальном аппарате ролевой теории элементов феноменологической программы одновременно позволяет и побуждает реконструировать ключевое для микросоциологии понятие (*ситуативного, эмерджентного*) *социального статуса* (социального «Я», социальной идентичности) с опорой на *беспредпосылочное* знание. Атомом — конституирующей ситуативные статусы участников — ситуации взаимодействия выступает веберовский «смысл», приведенный А. Шютцем к виду типизации. Чтобы актуализировалась выступающая фундаментом статуса социальная компетентность индивидов и состоялась ситуация взаимодействия, необходимо, чтобы один из участников определил своего партнера как носителя того или иного статуса. Эта базовая исследовательская посылка оказывается одновременно и единственно возможным атомарным *социологическим* описанием. В англоязычной литературе эта установка получила обозначение «in the eye of beholder».

Когнитивной основой для осуществления социальным деятелем операции типизирования, (например, в случае с гендерными статусами «отнесения к категории по полу»), выступает то, что можно назвать, развивая тезис Ч.Р. Миллса, «социологической невообразимостью» т.е. отсутствие у носителей обыденного сознания «социологического воображения». Материалом для типизирования выступает предъявляемое Другим статусное изображение. Типизировав своего партнера как носителя определенного статуса, индивид одновременно осуществляет комплементарную самотипизацию. Затем участники интеракции выбирают из конечного перечня опций содержание интеракции. Так, если речь идет о ректоре и студентах, то в качестве содержания может выступить (а) прием по личному вопросу, (б) выступление ректора на церемонии вручения дипломов, (в) поздравление ректора с юбилеем и др.

Социологическая невообразимость означает неспособность обыденного сознания преодолеть иллюзию целостности субъекта, прийти к осознанию автономного существования ролевых сущностей. Так, в повседневности мы склонны атрибутировать поведенческие проявления, перформансы не ситуативной (=временной и эфемерной) социальной функции человека, а самому человеку, его экзистенциальному «Я». И, соответственно, реагировать на него; например, обижаться на *человека*, а не на его ситуативную идентичность.

Продуктивным для анализа функционирования этого явления оказывается обращение к феноменологической «естественной установке». «Этот мир прост, утверждается сам собой и, кажется, не требует никакого дальнейшего объяснения. Мы обладаем по отношению к нему естественной установкой, которая на удивление беспроblemна. По словам Шютца, это была та мысль, которая более всего интересовала его в феноменологии Гуссерля. ... Жизненный мир есть невопрошаемая действительность»¹.

В данном случае речь идет о не-проблематизации видимости тела в его связке с психе́. В результате мы получаем третью эгологическую форму Гуссерля — понятие «Я», как оно используется в повседневном мышлении. Причем отличие от неодушевленных объектов здесь неочевидное: в формальном плане, исходя из «онтологической формы действительности», к которой принадлежит и «психе́», и вещь, они аналогичны. Именно материальный, телесный компонент «человека» обеспечивает нас видимостью, «очевидностью субъекта как уникального, незаменимого и идентичного самому себе»².

«Психе́ с самого начала воплощено..., оно как элемент природы (по)является в связи с материальной вещью — телом животного или человека»³. Об этом же писала, рассуждая о природе половой идентичности — частном случае проблемы онтологии социального статуса — С. де Бовуар: «Итак, я — это мое тело, во всяком случае, в меру моего опыта и наоборот мое тело — это нечто вроде естественного субъекта, предварительные наброски моего существования в целом»⁴.

Наблюдаемый социальным деятелем в каждом данном взаимодействии статус его партнера, будучи частью «Я» Другого, полагается неотделимым от него. Как справедливо отмечает И. Гофман, в процессе взаимодействия люди предполагают, что каждый участник процесса взаимодействия обладает «сущностной природой» («естеством»), распознаваемой через «естественные знаки, которые являются данностью или выражаются людьми»⁵.

Формулируя проблему в терминах феноменологической редукции, мы пытаемся выстроить индуцирующую статусные позиции социальную ситуацию «снизу». Здесь возникает закономерный вопрос о регулярности типизаций. Шютц задает этот вопрос применительно к

¹ Абельс Х. Романтика, феноменологическая социология и качественное социальное исследование // Журнал социологии и социальной антропологии. 1998. Т. 1. № 1. С. 125.

² Пешё М. Прописные истины: лингвистика, семантика, философия / Пер. с франц. Л.А. Илюшечкиной // Квадратура смысла: французская школа анализа дискурса / Пер. с фр. и португ.; общ. ред. и вступ. ст. П. Серию; предисл. Ю.С.Степанова. М.: Прогресс, 1999. С. 262.

³ Шютц А. Смысловая структура повседневного мира: очерки по феноменологической социологии / Сост. А.Я. Алхасов; пер. с англ. А.Я. Алхасова, Н.Я. Мазлумяновой; научн. ред. перевода Г.С. Батыгин. М.: Институт Фонда «Общественное мнение», 2003. С. 13.

⁴ Бовуар С. де. Второй пол: Пер. с фр. / Общ. ред. и вступ. статья С.Айвазовой. СПб.: Алетейя, 1997. С. 63.

⁵ Goffman E. Gender Display // Studies in the anthropology of visual communication. 1976. No. 3. P. 75.

своей собаке Роверу: «почему, — спрашивает он, — я склонен видеть в Ровере то млекопитающее, то собаку, то заболевшего Ровера?». И дает следующий ответ: структура объективированных типизаций не одномерна. Проблема «где остановиться в своем типизировании» решается акторами исходя из «наличной цели» (*purpose at hand*), которая «как раз и определяет элементы, которые являются релевантными по отношению к ней»¹.

«Вся типизация состоит в уравнивании черт, релевантных для конкретной непосредственной цели, ради которой образуется тип, и в пренебрежении теми индивидуальными отличиями типизируемых объектов, которые нерелевантны для такой цели. Не существует такой вещи как чистый или простой тип. Все типы являются реляционными терминами, имеющими, заимствуя выражение из математики подстрочный индекс, относящееся к цели, ради которой они созданы»².

Поскольку цели, преследуемые людьми в их уникальных «биографических» ситуациях недоступны нашему пониманию, мы должны ввести постулат об их институциональной природе. Так мы приходим к реалистичной установке: типы, не исходя от людей, они оказываются промежуточным образованием между институционально определяемой целью и индивидуальным действием. Следующий шаг здесь — признать цель элементом ситуации. Тогда очевидным становится противоречие: ситуация определяется индивидом на основе типизации, которая, в свою очередь, строится на основе ситуативно обусловленной цели.

Аналогичные рассуждения приводят позднего Гофмана к идее примата институциональных образцов: «Безусловно, определения ситуаций имеют реальные последствия, но они могут лишь очень незначительно повлиять на ход событий. ... Хотя люди почти всегда находят ... определения ситуации, они *не создают* эти определения. Скорее, их создает общество, а люди всего лишь прикидывают и оценивают, ... и затем действуют»³. Г.С. Батыгин приводит высказывание самого Гофмана: «Как афористически выразился Гофман, “дело не в интеракции, а во фрейме”»⁴.

Однако типизирование, определение ситуаций, ситуационный перформанс и *quality* (по Гофману) наблюдаются все же на человеческих единицах, хотя и с поправкой на «смерть автора». Как пишет в этой связи Э. Пучинелли Орланди, «[субъект] не считается ответственным за него [за смысл], хотя и мыслится как составная часть процесса производства смысла»⁵.

Поскольку большая часть социального действования носит речевой характер, актуализируется задача по реконструкции социального статуса, *quality* авторов по речевому перформансу. Эта задача выступает ключевой для французской школы анализа дискурса, а также для микро-социолингвистических штудий. Дискурс-аналитическая концепция приведения множества рассеянных высказываний к позиционному единству, полностью устраняющая из модели говорения человека оперирует перечнем «структурных элементов» условий порождения дискурса, содержащим импликации когнитивных процессов (см. таблицу).

М. Пешё также добавляет в модель переменную «образ референта/ситуации/контекста», разделяемый участниками взаимодействия, в формулировке «что говорится?». Несмотря на его конвенциональную онтологическую самостоятельность, он призван *точнее* определять позиции участников интеракции. Отметим также, что в свете предшествующих рассуждений построение $I_d(B)$ — конститутивную типизацию Другого vs самого себя — следовало бы пере-

¹ Шюц А. Структура повседневного мышления // Социологические исследования. 1988. № 2. С. 130.

² Шюц А. Смысловая структура повседневного мира: очерки по феноменологической социологии / Сост. А.Я. Алхасов; пер. с англ. А.Я. Алхасова, Н.Я. Мазлумяновой; научн. ред. перевода Г.С. Батыгин. М.: Институт Фонда «Общественное мнение», 2003. С.269.

³ Ковалев А.Д. Предисловие. «Фрейм-анализ» Гофмана с точки зрения переводчика // Социологическое обозрение. 2002. Т. 2. № 2. С. 12.

⁴ Батыгин Г.С. Континуум фреймов: социологическая теория Ирвинга Гофмана // Гофман И. Анализ фреймов: эссе об организации повседневного опыта / Пер. с англ. под ред. Г.С. Батыгина и Л.А. Козловой; вступ. ст. Г.С. Батыгина. М.: Институт социологии РАН, 2003. С. 33.

⁵ Пульчинелли Орланди Э. К вопросу о методе и объекте анализа дискурса / Пер. с португ. Б.П. Нарумова // Квадратура смысла: французская школа анализа дискурса / Пер. с фр. и португ.; общ. ред. и вступ. ст. П. Серию; предисл. Ю.С. Степанова. М.: Прогресс, 1999. С. 201.

местить в начало списка. Сам Пешё не мог этого не осознавать, вслед за Л. Альтюссером он пишет в «Прописных истинах», что «идентичность возникает из идентификации и обращения к субъекту»¹. Поэтому такой порядок взаимных и самотипизаций, по видимому, обусловлен чисто техническими моментами.

Таблица

Структурные элементы условий порождения дискурса²

Структурные элементы условия и ролевой диалог			
Символ, обозначающий воображаемое построение		Значение символа	Имплицитный вопрос, «ответ» на который лежит в основе соответствующего воображаемого построения
A	$I_A(A)$	Образ ³ позиции A для занимающего позицию A	«Кто я таков, чтобы ему это говорить?»
	$I_A(B)$	Образ позиции B для занимающего позицию A	«Кто он таков, что я ему это говорю?»
B	$I_B(B)$	Образ позиции B для занимающего позицию B	«Кто я таков, что он мне это говорит?»
	$I_B(A)$	Образ позиции A для занимающего позицию B	«Кто он таков, чтобы мне это говорить?»

Дискурс-аналитический проект направлен на поиск и систематизацию инвариантов (глубинных структур) сверхфразового уровня. Методологическая установка здесь не отличается от внутрилингвистической: инвариант есть абстрактная сущность, обнаруживающая себя в «почти бесконечном количестве "поверхностных представлений"»⁴. Если рассматривать минимальную поверхностную единицу дискурса — высказывание — как социальное действие, то речь идет о реконструкции ситуативного quality путем упорядочивания перформативных единиц.

Аналогичным образом формулируется одна из ключевых задач микросоциологической программы М. Халлидея. Он говорит о переключении языкового кода при переходе из одной социальной ситуации в другую. Тогда концептуализация ситуации (с выделением ее компонентов, включая ситуативные идентичности) становится коррелятом инварианта, а наблюдаемый эмпирически массив (разноуровневых⁵) языковых средств и правил после упорядочивания с применением статистических процедур представляет собой вариативный ряд. Халлидей параметризует социальные контексты, которые вместе с наблюдаемыми в них языковым разнообразием он называет *регистром*, с привлечением аналогичного представленному выше набору переменных. Также он приводит различия между *социолектом* и *регистром*, где последовательно проводится мысль о перманентности первых и ситуативной текучести вторых⁶.

В обеих концепциях инвариант мыслится в позиционно-семантических терминах⁷: основ-

¹ Пешё М. Прописные истины: лингвистика, семантика, философия / Пер. с франц. Л.А.Илющечкиной // Квадратура смысла: французская школа анализа дискурса / Пер. с фр. и португ.; общ. ред. и вступ. ст. П. Серио; предисл. Ю.С. Степанова. М.: Прогресс, 1999. С. 262.

² Пешё М. Контент-анализ и теория дискурса / Пер. с франц. И.Б.Иткина // Квадратура смысла: французская школа анализа дискурса / Пер. с фр. и португ.; общ. ред. и вступ. ст. П. Серио; предисл. Ю.С. Степанова. М.: Прогресс, 1999. С. 321.

³ Пешё делает многочисленные оговорки, что это именно *образ*, *идеальное отражение*, нежели *объективные сущности*. Между первыми и вторыми, по его мнению, должны существовать «правила проекции» (321). Этот «образ» и феноменологическое «типизирование» или интеракционистское «определение (ситуации)» здесь вполне взаимозаменяемы.

⁴ Пешё М. Контент-анализ и теория дискурса / Пер. с франц. И.Б.Иткина // Квадратура смысла: французская школа анализа дискурса / Пер. с фр. и португ.; общ. ред. и вступ. ст. П. Серио; предисл. Ю.С. Степанова. М.: Прогресс, 1999. С. 330

⁵ Поскольку речь идет о *социолингвистической* концепции, язык здесь не заканчивается на уровне предложений. Он здесь вообще нигде не начинается и не заканчивается, так как сосюрговская дихотомия полагается преодоленной.

⁶ Карасик В.И. Язык социального статуса. М.: Гнозис, 2002. С. 27–28.

⁷ М. Пешё пишет о «глобальной семантической эквивалентности», что эквивалентно инвариантности.

ной смысл сообщению придают не языковые единицы, а ситуация, в которой оно звучит. Итак, мы имеем два вида и два пласта переменных: социологические идентификационно-позиционные сущности и лингвистические перформансы. *Статус в речи*, таким образом, есть ситуативная социальная идентичность, эмпирически регистрируемая как делание, перформанс — здесь говорение и текст как его результат. Этот речевой перформанс *per se* необходимо далее расчленить на перформанс как индикатор *quality* и перформанс как перформанс — то, вариативность чего на основе *quality* предстоит изучить¹. Это более масштабная задача, чем то, что решается в рамках анализа дискурса: там осуществляется односторонняя реконструкция. Концепция регистра оказывается здесь более удовлетворительной, но лишь на уровне постановки вопроса *vs* предоставления теоретико-методологических оснований для его решения. В результате мы оказываемся перед необходимостью привлечения обыденных схем для создания научных описаний ситуаций. Так, изучая флирт, мы должны глядя на лежащие перед нами "скрипты" подслушанной ситуации, (т.е. на основе перформанса), сделать вывод о том, что это флирт, а коммуниканты — участники флирта (т.е. определить *quality*), а затем начать изучать различия во флиртовом перформансе с позиций определенного по этому же перформансу *quality*. Экспликативная чистота этой операции понимания вызывает серьезные сомнения.

Дискурс-анализ репрезентаций Кыргызстана в период политического кризиса

Иванова Инна Игоревна,

Фонд «Общественное мнение», НИУ ВШЭ

Задача, для решения которой используется диагностическая процедура, формулируется просто: кто есть кто, и что есть что?²

Сегодня сложно переоценить значение аналитических практик в социологии: социологическая экспертиза и диагностика являются необходимыми элементами принятия социально ориентированных управленческих решений. Определяя соотношение диагностической процедуры с понятием социологического измерения, мы разделяем подход, обозначенный Г.С. Батыгиным: «Если не сводить измерение к действию по прикладыванию линейки к предмету, измерение превращается в диагностику — переход от “видимостей” к тому, что существует “на самом деле”»³.

Летом–осенью 2010 г. Фондом «Общественное мнение» в рамках социологической экспертизы кризиса в Кыргызстане был проведен цикл социологических исследований в ряде постсоветских стран (Кыргызстан, Китай, Таджикистан, Узбекистан, Казахстан). В результате исследований был получен материал, зафиксировавший уникальное состояние общества: несмотря на внешний кризисный макроклимат и негативную оценку происходящего в стране, среди населения были выявлены надежда на лучшее и даже уверенность в «светлом будущем». В сложившихся условиях методы массовых опросов не могли дать достаточно полный материал для глубокого анализа причинно-следственных связей. Требовалось исследование латентных обоснований позиций респондентов. В качестве метода был выбран дискурс-анализ транскриптов глубинных интервью. В выборку вошли интервью с кыргызами, узбеками, казахами и китайцами, представляющими различные социально-демографические группы по возрасту, полу, образованию, роду занятий, месту проживания.

В данной работе проиллюстрирована схема применения дискурс-анализа, цель которого — систематизировать разрозненные маркеры мнений и настроений людей, задачи — выявить ключевые идеи информантов и связи между ними, сравнить обобщенные точки зрения представителей различных национальностей.

¹ На первом этапе задача «социологична» по своему содержанию, но в силу речевой природы перформанса «лингвистична» по форме. На втором этапе мы имеем дело с чисто лингвистическими моментами.

² Батыгин Г.С. Предисловие // Девятко И.Ф. Диагностическая процедура в социологии. Очерк истории и теории. М.: Наука, 1993. С. 4.

³ Там же.

Для достижения цели и решения задач была использована процедура дискурс-анализа, предложенная Г.В. Градосельской¹:

1) концептуализация: выделение концептов из речевых маркеров информантов (т.е. нахождение выраженных в различной словесной форме опорных идей, составляющих смысловую основу текста информантов);

2) структурирование: состыковка концептов, определение связей между ними;

3) интерпретация: выявление глубинных настроений информантов (мифов сознания), определение источников (факторов появления) связей между концептами.

Предложенная процедура вполне согласуется с «задачей дискурс-аналитиков — изучить связь между дискурсом и реальностью»²: анализ фокусируется не только на лингвистической, но и на семантической структуре означающего и на ее связи с социальным контекстом. Единицы в дискурс-анализе — не слова и словосочетания, а смысловые блоки и причинно-следственные связи между ними.

Для иллюстрации рассмотрим фрагменты транскриптов интервью, относящиеся к подтеме «Образ Кыргызстана» (общая тема интервью — сценарное развитие Кыргызстана).

В соответствии с описанной схемой, **на первом этапе** дискурс-анализа из речевых маркеров были выделены 4 основных концепта:

1) Кыргызстан — это Родина; речевые маркеры: родной дом, семья, дом, дети, жена, родители, земля кыргызов, земля наших предков, «моя земля», я здесь родился/лась и вырос/ла, любимая страна, «мы родились и живем здесь»;

2) Кыргызстан — это красивая природа; речевые маркеры: горы, солнце, тепло, чистое (ясное) небо, теплый песок, бурные реки, горные озера, Бишкек; преобладает Иссык-Куль как предмет главной гордости, олицетворение красоты и богатства природы страны: «Иссык-Куль — наше богатство (жемчужина, любимое озеро)»;

3) Кыргызстан — это государственность; речевые маркеры: разнообразные нации вместе, единство, флаг, герб, своя земля, независимость, свой гимн, единство, стабильность, независимость, республика, народ; один из преобладающих маркеров — первое лицо: с ним связывается дальнейшая судьба Кыргызстана как государства;

4) Кыргызстан — это бардак; речевые маркеры: нестабильность, бардак, коррупция, злость, холод.

Обычно процесс выделения речевых маркеров, а затем концептов сложно представить в виде четкого алгоритма, однако в конкретном случае простая формализация оказалась возможной из-за того, что анализируемые фрагменты транскриптов интервью имеют схожую структуру. Они представляют собой номинативные предложения, ответы на вопрос «Скажите, пожалуйста, какие три первых слова, или словосочетания, или образа, возникают у Вас, когда Вы слышите слово «Кыргызстан? Почему именно это приходит Вам на ум?» Называемые слова и словосочетания являются в данном случае речевыми маркерами, а общая семантика становится обоснованием для их объединения в смысловые группы — концепты. Пример:

М: Скажите, пожалуйста, если я Вам скажу слово Кыргызстан, какие образы в Вашей голове возникают, какие словосочетания?

А: Бескрайние джайло, высокие горы, чистые реки... солнце над головой...

М: Почему именно эти образы у вас возникли?

А: Потому что у меня в первую очередь Кыргызстан связан больше всего с природой, именно с образами гор.

Подобные цепочки логико-семантических связей выявлены и в других интервью, что позволило объединить группу речевых маркеров с близкой семантикой в концепт «природа».

На **втором этапе** (структурирования) из транскриптов выделялись связи между концептами (см. Рисунок). Процедура выявления связей между концептами в нашем случае имела двухступенчатый характер. На первой ступени из контекста выделялись пары маркеров, на-

¹ Взято из текста лекции Г.В. Градосельской «Дискурс-анализ в социологических исследованиях».

² Филипс Н., Харди С. Что такое дискурс-анализ? / Пер. Е. Кожемякина // Современный дискурс-анализ. 2009. Т. 1. Вып. 1. С. 49. Режим доступа: http://www.discourseanalysis.org/ada1_1.pdf

званных информантами в разных частях сложноподчиненных предложений (в основном и, в основном, придаточном причины). На второй ступени велся поиск концептов, к которым относятся названные маркеры. Были выявлены 3 связи.

1. Связь между концептами «государственность» и «бардак». Первый озвучивался зачастую скорее как желаемый, чем реальный (желание единства на фоне межэтнических распрей). Эту связь иллюстрирует, например, следующий фрагмент:

«Кыргызстан — это единство [маркер в составе концепта «государственность», потому что до этого времени особо не думали об этом, и после событий, которые происходят у нас в Кыргызстане, те же межэтнические войны, в голову сразу приходит это. Я хочу, чтобы в Кыргызстане было единство, независимо от вероисповедания, нации и всего прочего».

2. Связь между концептами «Родина» и «природа»:

«Мы живем в Кыргызстане и гордимся своим озером Иссык-Куль» [маркер «мы живем» относится к концепту «Родина», «озеро Иссык-Куль» — к концепту «природа»].

«Когда я слышу слово Кыргызстан, у меня возникают образы своей Родины, образы гор, наши уважаемые предки, наши любимые матери. Из всего этого и состоит наш народ».

3. Связь между концептами «Родина» и «государственность»:

«Первое, конечно, — Родина. Второе для меня и для наших людей — это страна, первое лицо. Третье, наверное, — мои родные, близкие люди, то, что меня окружает много хороших людей. Потому что раз они меня окружают, и мы тесно живем в этой стране, в государстве, и мы все связаны».

На **третьем этапе** (интерпретации) получившаяся структура концептов была сопоставлена со сложившейся ситуацией в кыргызском обществе. Основной вывод: не ориентируясь в обстановке и испытывая «информационный голод», люди воспринимают происходящее вокруг через призму собственных мифов о сильных сторонах Кыргызстана и желаемых сценариев его развития. Возникает эффект идеализации своей страны, надежды на светлое будущее, в частности, на «первое лицо», которое расставит все на свои места.

Сравнивая восприятие образа Кыргызстана у информантов разных национальностей, мы зафиксировали различия. Для узбеков одним из преобладающих является концепт «война», имеющий в сознании респондентов негативную окраску, напрямую связанную с межнациональными распрями. Если для кыргызов кризис — стимул для «веры в лучшее», то для узбеков — повод для осуждения. Здесь же оценка кыргызов как «разбойников» и «клентяев». Эту группу респондентов легко направить в русло нетерпимости. Другой концепт для узбеков — «близкая страна», «соседи». Основной мотив — сочувствие, ностальгия по временам добрососедства. Очевидна неопределенность позиции населения, которую легко склонить в сторону толерантности.

Первый из основных концептов в подтеме «Образ Кыргызстана» у казахов — «нестабильность». Однако, если у узбеков окраска сходного концепта негативная, то здесь недостаток информации и слухи вызывают не агрессию, а безразличие и настороженность, желание быть в стороне от проблем. Другой часто встречающийся концепт, такой же, как и у узбеков, — «соседи». Окраска определенно положительная, эмоциональная нагрузка «жалость», «сочувствие». Третий концепт, «зона отдыха», связан со вторым концептом. Цепочка: «мы соседи, ездим отдыхать друг к другу, зона отдыха в Кыргызстане — озеро Иссык-Куль».

В сознании китайцев «Образ Кыргызстана» состоит из 4 основных концептов. Преобладает «перспективы для бизнеса» (все информанты — предприниматели). Два другие — географический/культурологический («сосед Китая», «маленькое азиатское государство», «народные обычаи», «образ жизни народа») и «нестабильность». Последний концепт — «несамостоятельное государство»: «страна бывшего СССР», «страна не может справиться с экономическими и другими проблемами, по крайней мере, развитие происходит очень медленно». Все ассоциации носят утилитарный характер, основаны на опыте и фактах. Негатив в оценках отсутствует, даже кризис воспринимается с точки зрения последствий для бизнеса.

Помимо преимуществ дискурс-анализа (возможность фиксации причинно-следственных связей и происходящих изменений в сознании говорящего/пишущего, выявление факторов

этих изменений и связей, объединение точечных мнений в единую структуру, определение соотношения социального и индивидуального), важно отметить недостатки. На данном этапе развития процедура дискурс-анализа трудоемка и длительна при обработке больших массивов информации, также нельзя полностью исключить влияние личности исследователя. Важная особенность дискурс-анализа: его использование целесообразно перед проведением количественного исследования, в условиях, когда объект будущего незнаком или изменил свои основные свойства.

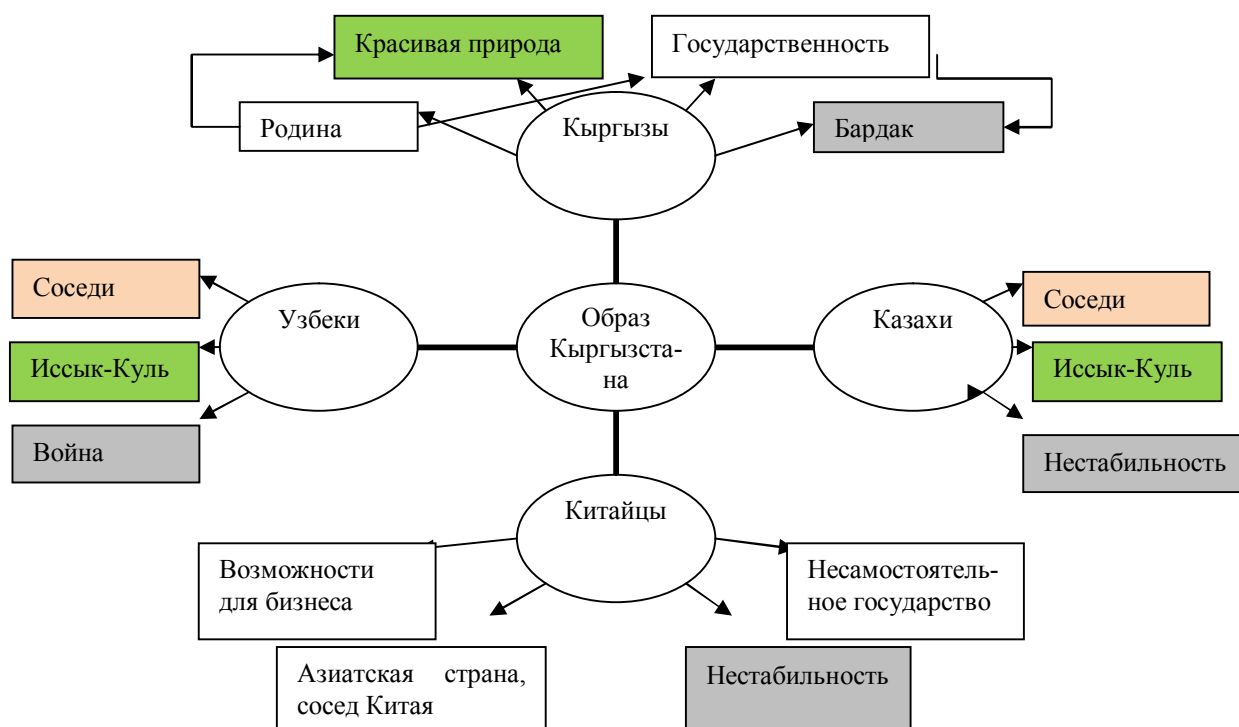


Рис. Образ Кыргызстана в восприятии кыргызов, узбеков, казахов, китайцев

В результате использования метода дискурс-анализа удалось объединить смысловые основы дискурсов отдельных людей в единую систему. В рамках подтемы интервью «Образ Кыргызстана» были определены установки информантов и их истоки, а также аргументы, способные модифицировать установки и направить поведение в определенное русло.

Анализ тримодальных данных на примере Интернет-сервисов социальных закладок

Игнатов Дмитрий Игоревич, *НИУ ВШЭ*
Магизов Руслан Азатович, *НИУ ВШЭ*

Введение

В работе представлен оригинальный метод трикластеризации, который использовался для поиска «плотных» троек вида (посетители, тэги, ресурсы) при анализе данных сервиса социальных закладок Bibsonomy. Предложенный нами метод является универсальным для анализа любых тримодальных данных, описываемых тернарными отношениями, и служит вычислительно эффективной альтернативой поиску трипонятий (в смысле анализа формальных понятий) и построению айсбергов трирешёток.

В настоящее время есть большое количество Интернет-сервисов, предлагающих совместное использование различных электронных ресурсов (social resource sharing systems), сгенерированных пользователями. Среди таких систем можно отметить Flickr (<http://www.flickr.com/>) — фотогалерея, delicio.us (<http://delicious.com/>) — сервис закладок, Bibsonomy (<http://www.bibsonomy.org/>) — сервис библиографических закладок; среди российских сервисов

социальных закладок можно упомянуть Bobrdobr.ru (<http://www.bobrdobr.ru/>) и Яндекс.Закладки (<http://zakladki.yandex.ru/>). Ключевая особенность таких сервисов в том, что пользователь «ставит закладку» на тот или иной ресурс, которая хранится в виде записи на удаленном сервере, а потому доступна с любого компьютера, вне зависимости от браузера. Стоит заметить, что современные веб-браузеры могут хранить закладки на сетевые ресурсы и поддерживать их синхронизацию через Интернет (например, Opera), но они не выполняют функцию совместного пользования. Помимо закладки пользователь такой системы может присвоить размещаемому им ресурсу (фотографии, Интернет-ссылке, библиографической записи) короткие ключевые слова, или тэги. С помощью таких тэгов пользователями независимо друг от друга осуществляется коллективная классификация ресурсов сервиса. Такие сервисы называют также коллективными системами использования тэгов (*Collaborative Tagging Systems*). Зная тэги, которые использует посетитель, можно находить пользователей, помечающих ресурсы такими же ключевыми словами, и рекомендовать такие ресурсы в качестве потенциально интересных, а самих пользователей — в качестве возможных друзей (людей с близкими вкусами). Сервис социальных закладок, снабженный таким рекомендательным механизмом, становится более полезным и приобретает статус рекомендательной системы на основе тэгов (несколько лет назад такое превращение произошло с сервисом *del.icio.us*, что повысило его популярность).

В основе социальных сервисов совместного использования ресурсов лежит структура данных, называемая фолксономией (*folksonomy*)¹. Фолксономия состоит из трёх множеств U , T , R — пользователей, ресурсов и тэгов, а также тернарного отношения Y между ними. Таким образом, к двум измерениям традиционной модели анализа потребительской корзины (покупки (*transactions*) и товары (*items*)) мы получили дополнительное измерение. И если в задаче анализа потребительской корзины отыскиваются (крупные) группы покупателей, приобретающие одинаковые продуктовые наборы, то при анализе фолксономий существенно находить группы пользователей, использующие одинаковые тэги для пометки некоторого множества ресурсов. Владельцы сервиса *Vibsonomy.org* используют модели и методы анализа формальных понятий (АФП)² для поиска таких групп (пользователи, ресурсы, тэги). Анализ формальных понятий является прикладной алгебраической дисциплиной, которая использует алгебраическую теорию решёток для формализации основной единицы человеческого мышления — понятия. Опираясь на объектно-признаковое представление данных, можно провести строгую формализацию понятия, как пары, состоящей из объема и содержания. Объем состоит из множества всех объектов, обладающих всеми признаками из содержания, а содержание — из множества всех общих признаков этих объектов; т.е. выполняется условие замкнутости, максимальности размера объема и содержания. Для формальных понятий также справедлив знаменитый закон обратного соотношения между размером объема и содержанием понятия. На множестве формальных понятий можно построить частичный порядок (иерархию) по отношению «быть более общим понятием», в котором понятие с большим по вложению объемом будет более общим для понятия с меньшим объемом. Данный частичный порядок обладает решётчными свойствами и называется решеткой понятий. Исходные объектно-признаковые данные получили название формального контекста в АФП, а для фолксономий (объекты – пользователи, признаки – тэги) исходные данные содержат одно дополнительное множество, которое в АФП принято называть условиями (для фолксономий это ресурсы). Существует так называемый триадический анализ формальных понятий³, расширение классической диадической модели, который имеет дело с трипонятиями. Такое трипонятие по-прежнему однородно и замкнуто, т.е. представляет собой максимальный параллелепипед в исходной таблице данных вида объекты-признаки-условия. Третья компонента понятия называется модусом. Существуют эффективные алгоритмы поиска формальных по-

¹ Jäschke R., Marinho L.B., Hotho A., Schmidt-Thieme L., Stumme G. Tag Recommendations in Folksonomies // Knowledge Discovery in Databases: PKDD 2007, 11th European Conference on Principles and Practice of Knowledge Discovery in Databases, Warsaw, Poland, September 17–21, 2007. 2007. P. 506–514.

² Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. Berlin et al.: Springer, 1999.

³ Lehmann F., Wille R. Triadic Approach to Formal Concept Analysis // Proceedings of the Third International Conference on Conceptual Structures: Applications, Implementation and Theory. Santa Cruz, 1995. P. 32–43.

ятий, на основе которых построены методы поиска трипонятий. Но в виду большой вычислительной сложности поиска формальных понятий, экспоненциальной в худшем случае от размера входа, методы поиска всех формальных трипонятий для реальных данных (несколько миллионов записей $\langle \text{пользователь, тег, ресурс} \rangle$), оказываются вычислительно неприемлемыми. Одним из традиционных решений является сокращения числа понятий, оставляемых для их дальнейшего анализа, на основе отбора по размеру объема или содержания. Таким путем пошли владельцы ресурса Bibsonomy.org. Нами ранее был предложен подход к поиску бикластеров¹, состоящих из двух множеств — объектов и признаков соответственно, — который оказался вычислительно более эффективным. Помимо этого, предложенное нами определение бикластера обладает полезным свойством: любое формальное понятие исходного контекста содержится в некотором бикластере этого контекста (в смысле покомпонентного вложения) при нулевом значении порога плотности. В этой работе мы обобщаем данный подход для случая триконтекстов (фолксономий), вводим определение (плотного) трикластера, предлагаем алгоритм их поиска и проводим эксперименты на данных сервиса Bibsonomy.org.

Математическая модель и алгоритм

Напомним определения диадического АФП. Формальным контекстом называется тройка $(G, M, I \subseteq G \times M)$, где G — множество объектов, M — множество признаков, а I — отношение инцидентности, показывающее, что данный объект g обладает признаком m (записывается gIm). Операторы Галуа сопоставляют подмножеству множества объектов множество всех их общих признаков или подмножеству множества признаков множество всех объектов, которые ими обладают. Пусть $A \subseteq G$, $B \subseteq M$, тогда операторы Галуа задаются следующими выражениями:

$$A' = \{m \mid m \in M \text{ и } (g, m) \in I\}$$

$$B' = \{g \mid g \in G \text{ и } (g, m) \in I\}.$$

Формальным понятием называется такая пара (A, B) , где $A \subseteq G$, $B \subseteq M$, что $A' = B$, а $B' = A$.

Дадим определение трипонятия триадического контекста (G, M, B, Y) как тройки вида (A_1, A_2, A_3) , максимальной в смысле покомпонентного вложения, где $A_1 \subseteq G$, $A_2 \subseteq M$, $A_3 \subseteq B$, такой что для $X_1 \subseteq G$, $X_2 \subseteq M$, $X_3 \subseteq B$ и $X_1 \times X_2 \times X_3 \subseteq Y$, из условия $A_1 \subseteq G$, $A_2 \subseteq M$, $A_3 \subseteq B$ всегда следует $(A_1, A_2, A_3) = (X_1, X_2, X_3)$. Как и в случае диадического анализа формальных понятий, мы будем пользоваться производными операторами (в диадическом случае это операторы Галуа) для удобства изложения и описания вычислений. Для краткости записи операторов обозначим триадический контекст несколько иначе (K_1, K_2, K_3, Y) . Пусть $\{i, j, k\} = \{1, 2, 3\}$, а $j < k$, тогда для $X \subseteq K_i$ и $Z \subseteq K_j \times K_k$, тогда (i) -производный оператор определяется следующими выражениями:

$$X \mapsto X^{(i)} = \{(a_j, a_i) \in K_j \times K_k \mid (a_i, a_j, a_k) \in Y \text{ для всех } a_i \in X\}$$

$$Z \mapsto Z^{(i)} = \{a_j \in K_i \mid (a_i, a_j, a_k) \in Y \text{ для всех } (a_j, a_k) \in Z\}.$$

В случае бикластеризации мы предложили определение бикластера как пары вида (m', g') для всех пар $(g, m) \in I$. Плотным бикластером мы назвали пару (m', g') , такую что плотность бикластера $\rho(m', g') = \frac{|m' \times g' \cap I|}{|m' \times g'|} \geq \rho_{\min}$, $\rho_{\min}$ — минимальное значение плотности, задаваемое пользователем.

В случае трикластеризации нам также понадобятся производные операторы, но только первого вида, сопоставляющие множеству элементов (например, объема) множества пар других элементов (первая компонента которых берётся из содержания, а вторая из модуса). Для упрощения нотации введем единое обозначение для такого оператора, действующего на трех различных множествах триконтекста. Мы будем применять этот оператор только к одноэлементным множествам, пусть $(g, m, b) \in I$, тогда

$$g' = \{(m, b) \in M \times B \mid (g, m, b) \in Y \text{ для всех } g \in G\}$$

¹ Игнатов Д.И., Каминская А.Ю., Кузнецов С.О., Магизов Р.А. Метод бикластеризации на основе объектных и признаковых замыканий // Интеллектуализация обработки информации: 8-я международная конференция. Сборник докладов. М.: МАКС Пресс, 2010. С. 140–143.

$$m' = \{(g, b) \in G \times B \mid (g, m, b) \in Y \text{ для всех } m \in M\}$$

$$b' = \{(g, m) \in G \times M \mid (g, m, b) \in Y \text{ для всех } b \in B\}.$$

Введём дополнительно бокс-операторы для элементов тройки $(g, m, b) \in I$:

$$g^\square = \{g \mid (g, b) \in m' \text{ или } (g, m) \in b'\}$$

$$m^\square = \{m \mid (m, b) \in g' \text{ или } (g, m) \in b'\}$$

$$b^\square = \{b \mid (m, b) \in g' \text{ или } (g, b) \in m'\}.$$

Трикластером назовем тройку вида $T = (g^\square, m^\square, b^\square)$, где $(g, m, b) \in I$. Плотным трикластером назовем трикластер T , плотность которого превышает минимальное значение — порог, заданный пользователем, т.е. $\rho(T) = |g^\square \times m^\square \times b^\square \cap I| / |g^\square \times m^\square \times b^\square| \geq \rho_{\min}$.

В таблице 1 приведён псевдокод предлагаемого нами алгоритма. Как и в двумерном случае, значение плотности трикластера лежит в интервале $(0, 1]$, а плотность трипонятия равна 1.

Таблица 1

Псевдокод алгоритма TRICL

TRICL((G,M,B,I) $\rho_{\min}$)
Вход: формальный триконтекст (G, M, B, I) и значение минимальной допустимой плотности трикластера.
Выход: TR — множество трикластеров.
For (g, m, b) in I $g' = \{(m, b) \in M \times B \mid (g, m, b) \in Y \text{ for all } g \in G\}$ $m' = \{(g, b) \in G \times B \mid (g, m, b) \in Y \text{ for all } m \in M\}$ $b' = \{(g, m) \in G \times M \mid (g, m, b) \in Y \text{ for all } b \in B\}$ For (g, m, b) in I $g^\square = \{g \mid (g, b) \in m' \text{ or } (g, m) \in b'\}$ $m^\square = \{m \mid (m, b) \in g' \text{ or } (g, m) \in b'\}$ $b^\square = \{b \mid (m, b) \in g' \text{ or } (g, b) \in m'\}$ if $\rho(g^\square, m^\square, b^\square) \geq \rho_{\min}$ then $T = (g^\square, m^\square, b^\square)$ $Tr.Add(T)$ Return Tr

Утверждение 1. Для некоторого формального триконтекста количество порождаемых (плотных) трикластеров не превышает $|I|$.

Утверждение 2. Время порождения всех трикластеров не превышает $O((|G||M|+|M||B|+|G||B|)|I|)$, а время порождения всех плотных трикластеров $O((|G||M|+|M||B|+|G||B|)|I||G||M||B|)$.

Что касается особенностей реализации алгоритма, то для избежания порождения дубликатов можно применять хеширование, т.е. для каждого нового трикластера находить значение хеш-функции, которое можно использовать в качестве ключа для хранения уникального трикластера в выходном массиве (словаре).

Гипотеза. Все формальные трипонятия исходного триконтекста (G, M, B, Y) содержатся в смысле покомпонентного вложения в некотором трикластере этого контекста.

Конечно, при увеличении минимального порога плотности до некоторого положительного значения данная гипотеза оказывается неверна. Однако формулировка гипотезы как положительного утверждения в настоящее время несколько поспешна и требует дополнительной проверки. У трикластера довольно простая геометрическая интерпретация — это параллелепипед в данных, в котором содержится «плотный» трехмерный крест, образованный пересечением трех меньших параллелепипедов с плотностью 1 в ячейке (g, m, b) .

Обсудим другие полезные свойства трикластеров. Рассмотрим триконтекст (Пользователи, Тэги, Книги) в таблице 2. Трирешетка понятий такого контекста содержит $3^3=27$ различных понятий, в то время как TRICL находит только один трикластер ($\{\text{Антон, Алёна, Настя}\}$,

{классика, судьба, чтиво}, {Золя.Западня, Достоевский.Крокодил, Островский.Гроза}) с плотностью $\rho=0,89$.

Таблица 2

«Слои» триконтекста (Пользователи, Тэги, Книги)

	классика	судьба	чтиво
Антон		X	X
Алёна	X	X	X
Настя	X	X	X
Золя. Западня			

	классика	судьба	чтиво
Антон	X	X	X
Алёна	X		X
Настя	X	X	X
Достоевский. Крокодил			

	классика	судьба	чтиво
Антон	X	X	X
Алёна	X	X	X
Настя	X	X	
Островский. Гроза			

В случае реальных данных вычисление всех трипонятий становится практически невозможным, да и вряд ли будет иметь для экспертов особый смысл анализировать вручную количество понятий, превышающее несколько тысяч. Поэтому такие плотные трикластеры при разумном выборе порога являются средством описания трехкомпонентных сообществ пользователей в данных фолксономий. Что касается условия отбора плотных трикластеров, то, к сожалению, оно не является ни монотонным, ни антимонотонным (в более плотном трикластере может содержаться менее плотный и наоборот).

Реальные данные и эксперименты

Для экспериментов мы выбрали реальные данные¹ сервиса Bibsonomy.org, содержащие 816197 записей вида (пользователь, тэг, ресурс), представленные в качестве конкурсного набора данных на международной конференции European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases в 2008 году рамках состязания Discovery Challenge. Фактически мы имеем триконтекст, которому соответствует параллелепипед размерами $|U||T||R|=2337 \times 67464 \times 28920$. В качестве экспериментального оборудования нами использовался компьютер с процессором Pentium-DualCore с частотой 2,5ГГц и с 2ГБ оперативной памяти. В качестве языка программирования мы выбрали высокоуровневый мультипарадигменный язык Python версии 2.7.1, идеально подходящий для быстрого прототипирования реализаций алгоритмов.

Первые эксперименты на 100 тысячах исходных записей, которые соответствуют параллелепипеду размером $59 \times 5823 \times 28920 = 9935668440$, показали, что алгоритм Tricl породил 4462 трикластера приблизительно за 3 часа (с учетом подсчета плотностей). Время вычисления всех трикластеров составило 1276 с, т.е. примерно 21 минуту (без учета подсчёта плотности). Как видим, отдельную проблему представляет собой точное вычисление плотности, которое сейчас осуществляется полной проверкой на принадлежность исходному триконтексту (отношению Y) тройки (g, m, b) — элемента конкретного трикластера T . Число таких проверок сопоставимо с размером параллелепипеда, соответствующего контексту. Альтернативным способом вычисления является рандомизированный приблизительный подсчет плотности, основанный на случайном выборе только N -ой доли ячеек трикластера для проверки принадлежности триконтексту. В случае хорошей корреляции этих приближенных значений с плотностью, такую оценку плотности можно использовать для отбора трикластеров по порогу её величины.

Интересным для исследователя является выполнение степенных законов распределения (Power Law) на данных бибсономии. Характерные примеры таких распределений наблюдаются

¹ <http://www.kde.cs.uni-kassel.de/ws/rsdc08/>

ся для посетителей сайтов, телефонных звонков, книг бестселлеров, интенсивности войн и т.п. Мы исследовали отдельно гистограммы распределения посетителей по количеству внесенных в систему пар (тэг, документ), тэгов — по количеству пар (пользователь, документ), ресурсов — по количеству пар (пользователь, тэг).

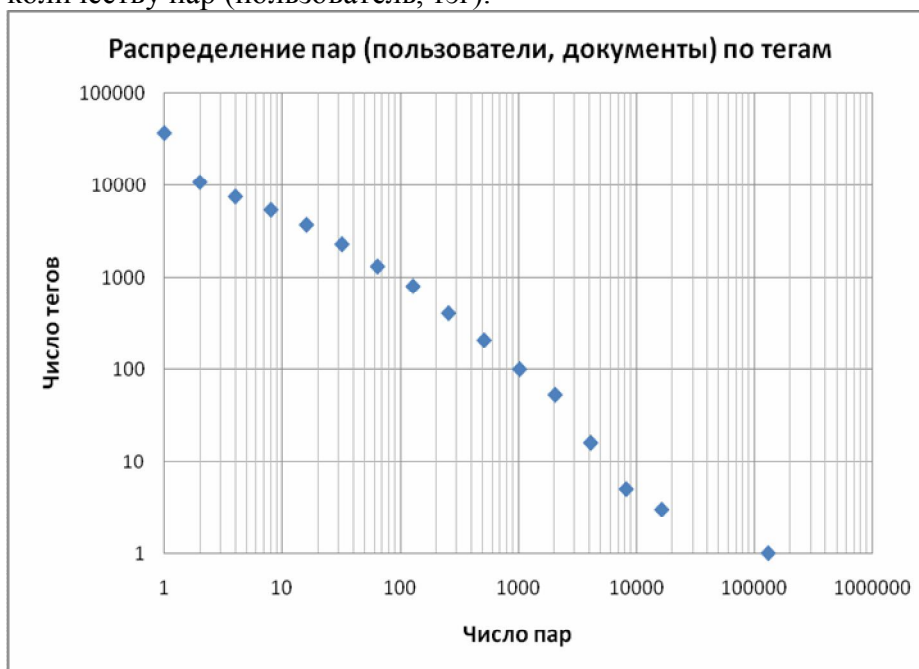


Рис. 1.

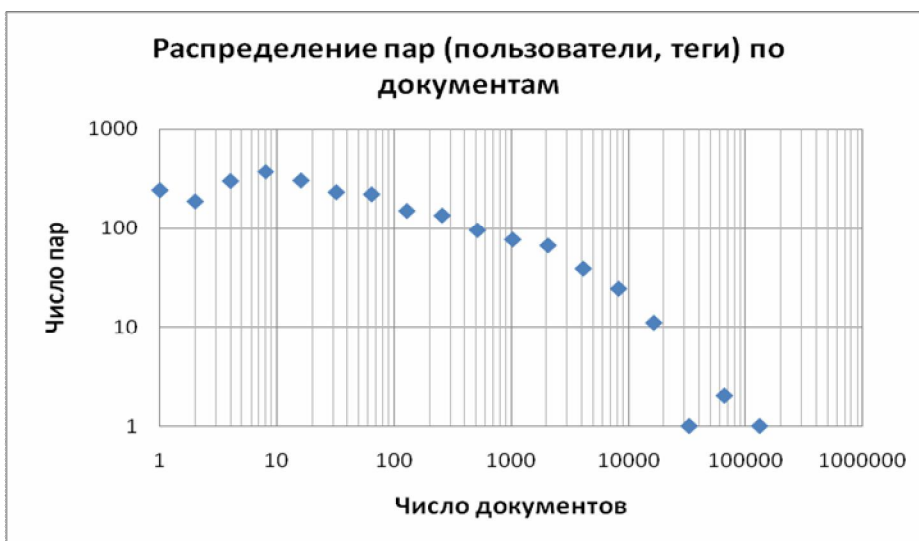


Рис. 2.

Для данных по количеству пар (пользователи, тэги) гипотеза о наличии степенного распределения не подтвердилась (рис. 2), хотя для распределения пар (пользователь, документ) этого нельзя с уверенностью сказать (рис. 1). Причиной этого могло послужить присутствие спамеров, которые генерировали высокочастотные тэги и документы. Распределение количества пар (пользователь, тэг), напротив демонстрирует степенной закон (явно выраженный прямолинейный участок в масштабе log-log, см рис. 3). Приблизительная оценка показателя a степени в законе распределения $p(x)=Cx^{-a}$ диаграммы на рис. 3 равна 1,5 (методика расчета изложена в работе Ньюмана¹). Как видим, первые четыре наиболее высокочастотных тэга (табл. 3) выполняют служебную функцию и, скорее всего, порождаются системой автоматически.

¹ Newman M.E.J. Power laws, Pareto distributions and Zipf's law // Contemporary Physics. 2005. Vol. 46. No. 5.

Для данных по количеству пар (пользователи, тэги) гипотеза о наличии степенного распределения не подтвердилась (рис. 2), хотя для распределения пар (пользователь, документ) этого нельзя с уверенностью сказать (рис. 1). Причиной этого могло послужить присутствие спамеров, которые генерировали высокочастотные тэги и документы. Распределение количества пар (пользователь, тэг), напротив демонстрирует степенной закон (явно выраженный прямолинейный участок в масштабе log-log, см рис. 3). Приблизительная оценка показателя a степени в законе распределения $p(x)=Cx^{-a}$ диаграммы на рис. 3 равна 1,5 (методика расчета изложена в работе Ньюмана¹).

Как видим, первые четыре наиболее высокочастотных тэга (табл. 3) выполняют служебную функцию и, скорее всего, порождаются системой автоматически.



Рис. 3.

Таблица 3

Топ10 + 1 тэг	
Тэг	Частота в парах присваиваний (пользователь, документ)
imported	66636
public	15666
system:imported	11294
nn	9147
video	7610
books	6214
software	5021
tools	4423
web2.0	4215
web	4071
blog	3439

Размеры трикластера — 2 пользователя, 27 тэгов и 8 документов, объём — 432, а плотность — 0.123. Даже такие, казалось бы, невысокие значения плотности, тем не менее, являются хорошим показателем релевантности трикластера, достаточно упомянуть, что плотность триконтекста, по которому был найден трикластер, составляет примерно 10^{-5} . Это позволяет нам с уверенностью говорить, что мы «нащупали сгусток плотности» в трикластере Т. Интерпретация трикластера Т довольно проста, мы видим все ресурсы (их идентификаторы) и тэги, которыми пользовались клиенты сервиса Bibsonomy с номерами 21 и 22 в некоторый временной промежуток, определяемый с момента запуска сервиса до появления 100000 записи.

¹ Newman M.E.J. Power laws, Pareto distributions and Zipf's law // Contemporary Physics. 2005. Vol. 46. No. 5.

Заключение

Результаты первых экспериментов позволят с уверенностью говорить, что метод находит релевантные трикластеры за приемлемое время (для сравнения, владельцы сайта сервиса Vibsonom.ru провели поиск крупных трипонятий на 600000 записях сервиса del.ici.ou.us примерно за 13 часов на мощном серверном оборудовании¹). Однако данный метод нуждается в дальнейшем тестировании и доработке, как с точки зрения оценки мер качества кластеризации, так и с точки зрения оптимизации вычислений. Исследование характера распределений числа использования пользователем тэга или ресурса также является предметом дальнейшего исследования, т.к. следование степенным законам вполне оправдывает жадные стратегии отбора трикластеров по величине плотности. Метод может быть полезен специалистам в области прикладной математики и социологии, работающих с социальными Интернет-сервисами, в особенности, с такими, в которых могут возникать тернарные отношения на исследуемых объектах и ресурсах, а также при анализе сообществ, возникающих в фолксономиях. Например, существует термин «biclique communities», определяющий множество членов сообщества через множество их общих интересов, а бикластер, как известно, является ослаблением понятия биклики и может выполнять сходную функцию при выявлении сообществ. И в нашем случае мы имеем похожую ситуацию, в которой от вычислительно более сложно порождаемого объекта — трипонятия — мы переходим к вычислительно более эффективному, но менее жёсткому — трикластеру.

Язык синтаксических структур и структурирование текста

Кобзарева Татьяна Юрьевна, РГГУ

В статье рассматриваются некоторые проблемы синтаксического анализа русского предложения и специфика их решения в модульной системе автоматического анализа MARS², где поверхностно-синтаксическому моделированию структуры ситуаций предшествует сегментация предложения. В свете быстро совершенствующихся возможностей виртуального хранения информации важнейшим источником при сборе и структурировании социологической информации являются сообщения СМИ и новостных лент, другая информация из сети Internet, отчеты коммерческих и общественных организаций и т.п. Объем исходных документов исчисляется десятками тысяч, а при обработке коротких сообщений, новостей и статей в СМИ — сотнями тысяч источников. Все более актуальными становятся задачи автоматического поиска информации. Лингвистика, и в первую очередь компьютерная лингвистика, уже более 50-ти лет занимается компьютерным анализом текста. Результаты работ по автоматическому анализу, позволяющие учитывать при поиске информации связи слов в тексте, очень важны для решения задач сбора и структурирования информации, однако в настоящее время они практически не используются.

Информация линейной структуры предложения

При автоматическом анализе исходным объектом является последовательность слов и знаков препинания текста — его означающее. Фердинанд де Соссюр, один из самых блестящих лингвистов XX века, отмечал, что означающее, «являясь по своей природе воспринимаемым на слух, развертывается во времени и характеризуется заимствованными у времени признаками: а) оно обладает протяженностью, и б) эта протяженность имеет одно измерение — это линия. Об этом совершенно очевидном принципе сплошь и рядом не упоминают вообще, по-видимому, именно потому, что считают его чересчур простым, между тем это весьма

¹ Jäschke R., Hotho A., Schmitz Ch., Ganter B., Stumme G. TRIAS - An Algorithm for Mining Iceberg Tri-Lattices // Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), 18-22 December 2006, Hong Kong, China. Hong Kong, 2006. P. 907–911.

² Система автоматического поверхностно-синтаксического анализа русского предложения MARS (modular analysis of Russian sentence), разрабатываемая автором в РГГУ при частичной поддержке РФФИ — грант № 09-06-00275-а.

существенный принцип и последствия его неисчислимы. <...> От него зависит весь механизм языка»¹.

Одномерному объекту, состоящему из слов и знаков препинания, мы хотим в конечном счете поставить в соответствие описание его означаемого — смысла, изображаемого в виде некого конструкта, представляющего смысловую — семантическую — структуру совокупности ситуаций, о которых идет речь в тексте. Информация о смысле текста может быть получена из двух источников. Один из них — это информация о его компонентах: — словах и знаках препинания, которую мы можем задать словарно, — и второй — информация линейного порядка следования компонент текста, для описания которой нужна специальная грамматика.

Минимальной компонентой моделирования смысла текста является ситуация². Для описания всей совокупности ситуаций текста нужно уметь найти и построить структуру каждой из них. Такую ситуацию описывает, например, простое предложение с одной предикативной вершиной, типа *Отец читает газету*.

Переход от линейного представления предложения к его смыслу, то есть к его семантической структуре, требует промежуточных этапов, как минимум, этапа синтаксического анализа, на котором мы извлекаем из предложения информацию о его синтаксической организации. Для представления этой информации, как и для описания смысла, нужен специальный язык.

Исходной единицей при синтаксическом анализе линейной структуры текста является предложение. Предложение состоит из слов и знаков препинания, идущих в определенном *неслучайном* порядке. Синтаксические и/или семантические возможности слов или словосочетаний, функционирующих как одно слово, в системах автоматического анализа содержатся в словарях. Каждому слову мы можем задать его потенциальные синтаксические возможности, определяющие его участие в *отдельно взятых ситуациях*. Основой построения семантического представления смысла предложения является синтаксическая структура совокупности ситуаций, которые содержатся в предложении.

Идеальная линейная структура русского предложения представляет собою соединение в одномерном пространстве предложения неограниченного числа ситуаций. Соответственно, его интерпретация чревата множеством потенциальных неоднозначностей разного рода.

Как облегчить на этапе синтаксического анализа построение соответствия идеальной линейной организации предложения, включающего в идеале неограниченное число ситуаций, его синтаксической структуре? В представляемой работе для этого используется грамматика, позволяющая перед синтаксической интерпретацией ситуаций разделить предложение на зоны, в которых эти ситуации локализуются.

Эта грамматика опирается на свойства линейной структуры предложения, позволяющие на всех этапах анализа систематически использовать обычно учитываемую лишь *ad hoc* информацию, которую несет линейный порядок слов и знаков препинания, в первую очередь, чрезвычайно семантилизованная система русской пунктуации. Ниже рассматриваются принципы организации системы автоматического синтаксического анализа на основе такой грамматики. В этой системе моделированию синтаксической структуры ситуаций предшествует деление предложения на части, в которых размещаются отдельные ситуации.

Язык синтаксических структур

Синтаксический анализ — необходимый мостик для перехода от одномерной структуры предложения к ее смыслу. В ходе синтаксического анализа моделируется синтаксическая структура предложения — то, что мы будем называть синтаксическим представлением линейной структуры. Это этап компьютерного анализа, когда мы находим в предложении подлежащее, сказуемое, прямое и косвенные дополнения, обстоятельства места, времени и т.д. Основополагающую роль для автоматического синтаксического анализа сыграли идеи структурного синтаксиса французского лингвиста Люсьена Теньера.

¹ Соссюр Ф де. Курс общей лингвистики // Труды по языкознанию. М.: Прогресс 1977. С. 103.

² Леонтьева Н.Н. Статус и возможности слабоструктурированного семантического пространства текста // Материалы V научно-практической конференции «Социологические методы в современной исследовательской практике». 2011. С. 334–343.

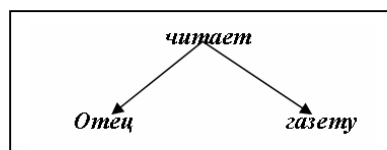
По Теньеру, «глагольный узел, который является центром предложения в большинстве европейских языков, выражает своего рода маленькую драму. Действительно, как в какой-нибудь драме, в нем обязательно имеется действие, а чаще всего также действующие лица и обстоятельства»¹. Действующие лица ситуации, вслед за Теньером, принято называть актантами, а обстоятельства — сирконстантами.

Модель ситуации, предложенная Теньером, позволила разработать универсальные способы описания структуры ситуаций, возможных ее участниках и их ролей в ситуации, и на этой основе описывать способы синтаксической манифестации этих ролей.

Моделирование синтаксической структуры каждой отдельно взятой ситуации опирается на словарную информацию о потенциальной синтаксической структуре ситуаций с вершиной — определенным предикатом, и о морфологических, синтаксических и семантических категориях слов, определяющих их способность участвовать в этих ситуациях в определенных ролях.

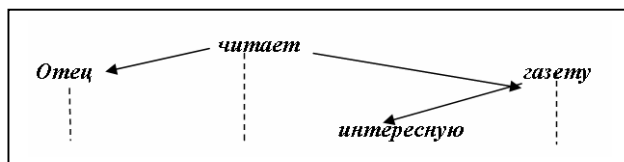
Вершиной в простом предложении чаще всего является глагольное сказуемое. Так, в предложении **Отец читает газету** сказуемое **читает** определяет структуру ситуации в целом, и из потенциальных синтаксических возможностей этого глагола следует, что слово **отец** для данной формы глагола является подлежащим (то есть на семантическом уровне называет субъект действия), а слово **газету** — его прямое дополнение (то есть его объект на семантическом уровне).

Теньер предложил представлять синтаксическую структуру простого предложения в виде *особого* дерева связей слов, которое *сохраняет по горизонтали порядок слов исходного предложения*. Такой способ представления заложил основу языка описания синтаксической структуры предложения, удобного для синтаксического представления русского предложения. Так, синтаксическое дерево связей предложения **Отец читает газету** будет по Теньеру таким:



Это дерево принято называть деревом *синтаксических зависимостей*. В данном случае оно состоит из трех *узлов* — слов, связанных *нумерованными отношениями подчинения*, отношениями, направленными от слова-хозяина к слову-слуге.

Такое представление синтаксической структуры дало возможность Теньеру открыть *свойство проективности* синтаксических структур, непосредственно обусловленное порядком слов в предложении и являющееся по Теньеру условием правильности структуры: если из узлов дерева Теньера опустить перпендикуляры, то в проективной структуре они не будут пересекаться со стрелками связей слов:

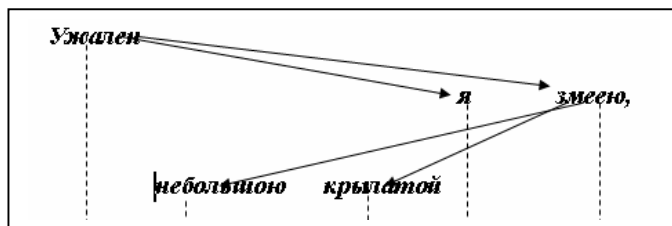


Таким образом, модель ситуации Теньера и его идея представлять синтаксическую структуру ситуации в виде дерева, а также использование понятия валентности для описания способности слова быть хозяином других слов, заложили основы языка, позволяющего описывать синтаксические значения предложения. Именно этот язык уже много лет служит продуктивным базисом исследования не только синтаксических, но и смысловых значений предложения, а в нарратологии — организации текста.

При синтаксическом анализе свойство проективности, описать которое стало возможно благодаря деревьям, предложенным Теньером, служит критерием правильности структуры,

¹ Теньер Л. Основы структурного синтаксиса. М.: Прогресс, 1988. С. 117.

если возникает необходимость выбора из гипотетически возможных структур¹, в подавляющем большинстве прозаических текстов любых жанров. Непроективность — редкое явление преимущественно разговорной или поэтической речи, например, непроективны строки Державина: *Ужален небольшою \ Крылатой я змеєю*:



Возможные в языке *синтаксические* связи слов задаются *списком синтагм*, описывающим морфологические характеристики слов, связанных определенными отношениями. Но определяющим всегда является смысл. В случаях, когда для различения связей, разных по смыслу, морфологических характеристик оказывается недостаточно, вводятся еще и семантические классы слов, дающие возможность правильно проинтерпретировать предложение.

Например, в предложении *Отец читает газету весь день* оба слова *газету* и *день* — существительные в винительном падеже, но мы понимаем, что прямое дополнение *газету* называет объект действия, а *день* — обстоятельство времени. Чтобы задать правило, следуя которому программа могла бы идентифицировать связи слов в парах *читает газету* и *читает день* как разные синтагмы, в словаре достаточно ввести семантический класс существительных ‘единица измерения времени’ и для синтагм, где существительное в винительном падеже будет обстоятельством времени, оговорить, что существительное должно относиться к этому классу.

Проблемы структурирования линейной организации предложения

Как мы уже отметили, даже в самых совершенных словарях, описывающих все нюансы синтаксических возможностей слов, мы можем описать только возможности слова входить в отдельно взятые простые ситуации. На деле же даже в простых предложениях мы сталкиваемся с более сложными видами линейной структуры. Их сложность, в частности, в том, что в них содержится несколько «маленьких драм» Теньера.

Например, в простом предложении (S1) *Проплыла замороженная в голубом стакане ярко-зеленая хвойная ветка* (О. Мандельштам) у существительного — субъекта действия есть три согласованных определения, одному из которых — *замороженная* — подчинена предложная группа *в голубом стакане*. Мы видим, что появление согласованного определения, выраженного причастием (но не только причастием), как бы «втягивает» в линейную структуру предложения свою ситуацию, т.е. свою «маленькую драму» со своими действующими лицами и обстоятельствами. При этом атрибутивные именные группы и предложные группы образуют **проективные рекурсивные** структуры, «матрешки», линейные структуры, состоящие из объектов, вложенных в им подобные. При анализе синтаксической структуры ситуации, вершина которой — сказуемое, эти многоярусные проективные рекурсивные структуры являются именами актантов и сирконстантов моделируемой ситуации.

Таким образом, даже в линейном пространстве простого предложения могут одновременно соединяться несколько простых ситуаций.

Кроме того, важной особенностью естественного языка является то, что язык не терпит однообразия и монотонности. Если бы мы говорили только простыми предложениями, состоящими из однозначных слов (что было бы чрезвычайно удобно с точки зрения автоматического анализа), получился бы очень скучный текст. Чтобы этого избежать, в языке существует множество разных механизмов. Один из них — построение внутри простых предложений матрешек атрибутивных определений.

¹ Лесерф Д. Применение программы и модели конфликтной ситуации к автоматическому синтаксическому анализу. М.: Научно-техническая информация, 1963, вып. 10.

Синтаксический анализ осложняют и другие механизмы, позволяющие соединять в одном предложении «маленькие драмы» Теньера и порождающие при этом многоярусные рекурсивные синтаксические структуры.

Чтобы избежать словесных и синтаксических структурных повторов, язык использует множество разных приемов. Таких, например, как трансформацию нескольких простых предложений в одно предложение с сочинением. Вместо того, чтобы сказать: *Отец очень устал, отец пошел домой*, — мы можем сказать: *Отец очень устал и пошел домой*. Или же мы можем превратить одно из этих предложений в деепричастный оборот: *Отец, очень устав, пошел домой*. Или же одно из них можно сделать придаточным с заменой какого-то слова во втором по порядку следования на местоимение: *Когда отец очень устал, он пошел домой*.

Части сложных предложений, являющиеся трансформациями простых и требующие по правилам русской пунктуации обособления знаками препинания, будем называть **сегментами**.

Перечисленные механизмы соединения простых предложений в единое сложное целое порождают линейные рекурсивные структуры (матрешки). Сегментная матрешка потенциально может представлять собою многоярусную комбинацию вложений сегментов в сегменты. В результате, линейная структура предложения представляет собою проекцию многоярусной системы рекурсивных структур, каждая из которых является синтаксической манифестацией некоторой ситуации, на одномерное пространство предложения. При описании в словаре синтаксических возможностей отдельных лексем мы описываем их способность участвовать в отдельно взятых ситуациях.

Чтобы оптимально использовать информацию словаря о синтаксических (или семантических) возможностях отдельных слов участвовать в ситуациях, надо произвести его делинеаризацию, т.е. найти линейные зоны, в которых локализуются отдельные ситуации.

Сегментная структура многосегментных предложений представляет собою, как и линейная структура атрибутивных именных групп, рекурсивную проективную структуру. Главная проблема построения сегментов связана с рекурсивностью линейной структуры, а именно с тем, что в простые предложения могут вкладываться разрывающие их сегменты, которые и сами могут быть разорваны вложениями других сегментов. Кусочки разорванных вложениями сегментов могут оказаться далеко друг от друга. Например, в предложении В. Набокова (S2) *Но, пожалуй, самым мучительным для него было то, что, догадываясь, какие идут о нем пересуды, сам он их не слышал*. Придаточное разорвано вложением в него деепричастного оборота, оборот — вложением в него придаточного. Построить сегменты, то есть найти их границы и соединить при этом эти кусочки и есть задача сегментации.

Главное свойство линейной структуры русского предложения, позволяющее решить эту задачу, — это *проективность сегментной структуры многосегментного предложения*. Оно заключается в том, что *никакая часть сегмента, разорванного вложением другого сегмента, не может оказаться внутри вложенного сегмента*.

Для сегментации была построена иерархия модулей анализа русского предложения¹, впервые реализующая на основе специальной грамматики сегментной структуры процедуры сегментации — построения частей предложения, границы которых по правилам русской пунктуации задаются знаками препинания. Порядок и задачи модулей были следующими:

Модуль	Задачи модуля
1. Постморфология	Разрешение несловарных проблем морфанализа
2. Разрешение омонимии частей речи	Разрешение морфологической омонимии частей речи
3. Предсегментация	Построение компонент текста для сегментации (проективных фрагментов именных и предложных групп, сложных сказуемых и т.д.)
4. Сегментация	Построение сегментов — частей предложения, эксплицитно заданных правилами пунктуации (придаточных предложений, деепричастных, определительных и других обособляемых оборотов)
5. Внутрисегментный анализ	Поиск связей слов внутри построенных сегментов
6. Межсегментный анализ	Построение связей сегментов

¹ Кобзарева Т.Ю. Иерархия задач поверхностно-синтаксического анализа русского предложения // НТИ. Сер. 2. 2007. № 1. С. 23–35.

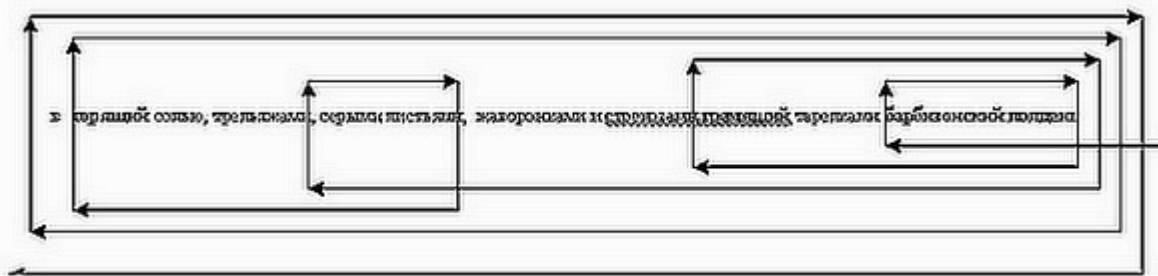
Рекурсивные процедуры анализа рекурсивных проективных структур

Система в целом и алгоритмы ее модулей организованы по принципу *divide et impera* — *разделяй и властвуй*, который оказался продуктивным и для общей организации системы, и при анализе сложных рекурсивных структур на всех уровнях синтаксического анализа русского предложения. Такой подход создает принципиальную возможность использовать модули системы для решения локальных задач, не требующих полного синтаксического анализа.

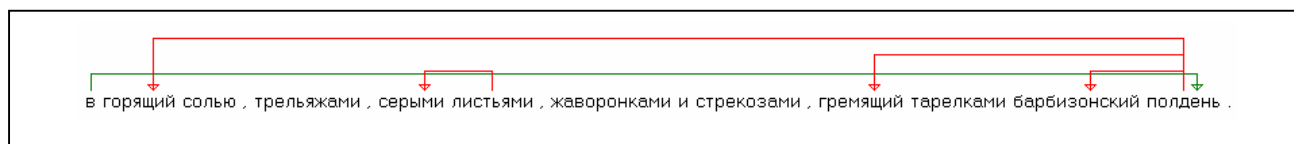
Так, модуль «Постморфология» строит морфологические характеристики слов и словосочетаний, которые не могут быть идентифицированы по словарю: имена собственные, многосоставные числительные в буквенной, цифровой и цифро-буквенной записи, окказиональные сокращения и т.д. Модуль «Разрешение омонимии частей речи» задуман как универсальная для русского языка программа, которая снимает все возможные типы морфологической омонимии частей речи по морфологическим характеристикам контекста¹. Модуль «Предсегментация» строит не только атрибутивные именные и предложные группы², как в приведенных примерах, но и сложные рекурсивные конструкции с именами собственными (*лидер «Нового союза» А. Паулаускас, обозреватель газеты «Новости недели» Аскар Умаров, глава фирмы «Монолитспортстройобъект» Иван Шебалдин*), конструкции с числительными (*16 тыс. химических авиабомб, 11 сентября 2001 года*), сложные сказуемые и т.д. Все перечисленные конструкции на этапе сегментации текста рассматриваются как единицы его линейной структуры.

Линейная структура предложения на всех ее уровнях рекурсивна. То есть мы имеем дело с вложениями некоторых структур в им подобные, причем внутренняя организация каждой вложенной структуры не значима для структуры, в которую она вложена. Поэтому на всех этапах анализа мы можем использовать рекурсивные процедуры, позволяющие чрезвычайно упростить грамматику каждого из этапов анализа и одновременно иметь возможность анализировать неограниченное число допустимых языком комбинаторных вариантов. Например, при построении проективного фрагмента атрибутивных именных групп анализ идет *справа налево* и для каждого очередного согласованного определения (прилагательного или причастия) ищется существительное-хозяин, а для предлога — существительное-слуга.

(S3) *Воздух, уксус и солнце уминались с зелеными тряпками в горящей солью, трельяжами, серыми листьями, жаворонками и стрекозами, гремющий тарелками барбизонский полдень.* (О. Мандельштам)



В результате строятся связи, определяющие линейную структуру вложений соответствующей матрешки предложной группы *в ... полдень*³:



¹ Кобзарева Т.Ю., Афанасьев Р.Н. Универсальный модуль предсинтаксического анализа омонимии частей речи в русском языке на основе словаря диагностических ситуаций // Компьютерная лингвистика и интеллектуальные технологии. Труды международного семинара Диалог'2002. Т. 2. М.: Наука, 2002. С. 258–268.

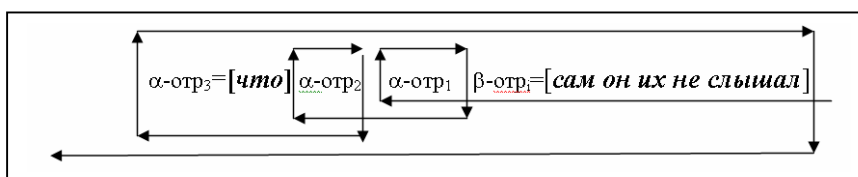
² Кобзарева Т.Ю. Некоторые свойства линейной структуры именных и предложных групп (Поверхностно-синтаксический анализ русского предложения) // Вестник РГГУ. № 8/07, Серия «Языкознание». 2007. С. 113–130.

³ Программная реализация модуля Предсегментация выполнена в рамках дипломной работы студентом ф-та прикладной математики МАИ П.В. Пашковским.

Рекурсивные процедуры оказываются чрезвычайно удобными и при построении сегментов¹. Анализ, так же как и при построении проективных фрагментов именных групп, ведется от конца предложения к его началу. При этом опорными элементами анализа являются α -отрезки — отрезки предложения между двумя знаками препинания, в которых есть слова, маркирующие их как минимальные компоненты сегмента: подчинительные союзы, деепричастные обороты, причастия или прилагательные, оставшиеся без хозяина после предсегментации и т.д. В результате такого порядка действий каждая матрешка анализируется, начиная с самого глубокого вложения. Каждая проанализированная вложенная структура элиминируется из зоны анализа, и, соответственно, анализ возвращается к исходной ситуации, что позволяет на основе минимального грамматического описания идентифицировать любые возможные контекстные комбинации сегментов.

Рассмотрим условные схемы построения сегментной матрешки при сегментации S4 и S5.

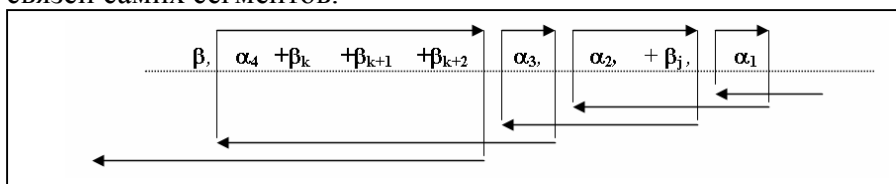
S4. $\beta\text{-отр}=[\text{Ho}]$, $\beta\text{-отр}_n=[\text{пожалуй}]$, $\beta\text{-отр}_m=[\text{самым мучительным для него было то}]$, $\alpha\text{-отр}_3=[\text{что}]$, $\alpha\text{-отр}_2=[\text{догадываясь}]$, $\alpha\text{-отр}_1=[\text{какие идут о нем пересуды}]$, $\beta\text{-отр}_1=[\text{сам он их не слышал}]$. (В. Набоков):



Для каждого очередного **справа налево** α -отрезка мы по правилам грамматики этого этапа анализа пытаемся присоединить к нему ближайший к нему справа β -отрезок. При таком порядке построения сегментов мы, как уже говорилось, начинаем анализ каждой сегментной матрешки с самого глубокого вложения. Процедура присоединения β -отрезков повторяется до того момента, пока очередной β -отрезок присоединить не удастся. При этом сегмент объявляется построенным и временно элиминируется из зоны анализа, что возвращает нас к исходной грамматической ситуации.

S5. $\beta n=[\text{Я знал}]$, $\alpha_4=[\text{что с ним заодно я убивал и созданный им мир}]$, $\beta_k=[\text{всю глупость}]$, $\beta_{k+1}=[\text{трусость}]$, $\beta_{k+2}=[\text{жестокость этого мира}]$, $\alpha_3=[\text{который с ним сросся во мне}]$, $\alpha_2=[\text{вытесняя до любого солнечного пейзажа}]$, $\beta_j=[\text{до детского воспоминания все сокровища}]$, $\alpha_1=[\text{собранные мною}]$. (В. Набоков)

Ниже изображена условная схема построения сегментов в примере S5, где при построении придаточного и деепричастного оборота β -отрезки присоединяются, так как при анализе мы находим сочинительные конструкции. После построения сегментов моделируется их внутренняя структура, то есть строится граф связей слов внутри построенных сегментов, а затем структура связей самих сегментов.



Заключение

Представлен подход к моделированию синтаксической структуры русского предложения, опирающийся на свойства проективности и рекурсивности его линейной структуры. Эти свойства позволяют сегментировать предложение перед моделированием составляющих его ситуаций, задаваемых в его линейной структуре иерархией рекурсивных вложений. Такой подход позволяет разделить процедуру анализа на лингвистически обоснованные ступени, построить универсальную грамматику для каждой из них и тем самым создает базис решения многих проблем автоматического анализа и поиска информации.

¹ Кобзарева Т.Ю. Принципы сегментационного анализа русского предложения // Московский лингвистический журнал. М.: Изд-во РГГУ, 2004. Т. 8. №1. С. 31–80.

Результаты работы каждого из модулей, предшествующих сегментации, и самой сегментации могут быть использованы не только для моделирования внутренней синтаксической структуры сегментов и построения графа связей сегментов. Они полезны и для семантического анализа, так как уточняют и дополняют морфологический анализ, необходимый для решения любых задач компьютерной обработки текста, и определяют зоны предложения, в которых находятся компоненты «маленьких драм» Теньера, максимально приближая результаты к следующему этапу — семантической интерпретации.

Предлагаемая грамматика синтаксического анализа позволяет решать и локальные задачи, требующие частичного структурирования входной текстовой информации.

Контент-анализ слухов в сети Интернет как элемент мониторинга репутации государственных учреждений и коммерческих компаний

Куликов Евгений Михайлович,
Куликова Наталья Александровна,
Кубанский государственный университет

В современном информационном обществе важное место занимает глобальная сеть Интернет. С каждым годом увеличивается численность ее аудитории, упрощаются формы и способы подключения к Интернету. В результате прогресса коммуникативных технологий нивелируются пространственно-временные границы. Если прежде масштаб деятельности большинства традиционных СМИ определялся размерами страны, региона, населенного пункта, то сейчас аудитория сети Интернет расширилась до планетарных масштабов¹.

Истоки возникновения Интернета лежат в разработках пакетной коммутации и открытой сетевой архитектуры, что характеризует его как продукт техногенного развития, предназначенный на начальном этапе для узкоспециализированных специалистов, каковыми и были пользователи первых компьютерных сетей. Однако на сегодняшний день упоминание термина «Интернет» может содержать в себе смыслы не только компьютерной, но и социальной направленности. Международная компьютерная сеть стала новым СМИ, общедоступным средством массовой коммуникации, площадкой экономических и политических операций, местом проведения досуга и даже религиозных действий. Из специализированной компьютерной сети Интернет за короткий срок приобрел черты социальной системы. Подобно тому, как Земля населена миллионами людей, Интернет «населяют» миллионы компьютеров, которые благодаря использованию единого протокола обмена информацией, особой системе адресации (IP-адреса, доменные имена) идентифицируют и находят друг друга. Взаимодействие в сети Интернет привело к появлению множества «виртуальных миров», субъектами которых являются реальные и виртуальные личности, группы и общности².

Фактически, создание Интернета привело к образованию виртуальной социальной реальности, где индивиды обучаются, общаются и «живут», как в обыденной реальной жизни, или зачастую подменяя ее. Причина этого в том, что всякая техническая новация влияет на социум, являясь одновременно и порождением научно-технического прогресса и социальным заказом. В межличностных коммуникациях по Интернету должна присутствовать база, необходимая для процесса взаимодействия индивидов и групп:

- аудитория ресурса, обладающая определенными общими ценностными ориентациями;
- социальная значимость информации с точки зрения возникновения и развития межличностного взаимодействия акторов;
- техническая среда, т.е. соответствующие средства, поддерживающие процесс взаимодействия.³

¹ Путилова Е.А. Интернет как фактор формирования информационного общества. Автореф. дис... канд. соц. наук. Тюмень, 2004.

² Силаева В.Л. Интернет как социальный феномен // Социологические исследования. 2008. № 11. С. 101–102.

³ Черемисин А.Г. Конструирование новой социальной и образовательной реальности на основе интернет-технологий. Автореф. дис...канд. М., 2009. С. 9–10.

В настоящее время значительно возросла скорость трансляции информации, однако, одновременно произошло и ухудшение ее качества. В сети Интернет коммуникация часто выглядит неперсонализированной, обезличенной, соответственно, у части пользователей появляется соблазн передачи сообщений, носящих недостаточно достоверный, а иногда и откровенно ложный характер. При этом возможности государства и общества в сфере контроля над сетью Интернет серьезно ограничены, учитывая планетарный характер этого СМК.

Это еще раз доказали события октября 2008 г., когда российские банки столкнулись с ажиотажем среди населения, стремящегося забрать свои вклады. Эта паника носила отнюдь не случайный характер, провокационная информация о банкротстве того или иного банка периодически появлялась в Интернете, муссировалась на сайтах, форумах, в чатах. Банки в различных регионах России констатировали рост количества желающих досрочно закрыть свои счета и пытались убедить своих клиентов, что для паники нет оснований. Между тем Россия оказалась охваченной слухами о мнимой потере ликвидности и банкротстве банковских учреждений. Причем менялась лишь география и наименование банка, а тип тревожного слухового сообщения был универсален и выглядел примерно так: «Банк "Х" разорился, срочно снимайте свои сбережения!» Судя по данным СМИ, тревожные слухи молниеносно охватили территорию РФ от Калининграда до Владивостока¹.

Однако было бы неправомерным утверждать, что целью агрессивных коммуникаторов был какой-то один или два банка по всей РФ. И если бы тревожные слухи касались бы какого-то одного банка «Х» в одном или нескольких субъектах РФ, можно было с уверенностью говорить либо о злонамеренных действиях конкурентов этого банка, либо о реальных проблемах в его деятельности. Однако когда речь идет о множестве не связанных между собой (не родственных) банках, а география слухов простирается с Запада на Восток РФ при сохранении временных параметров начала–середины октября 2008 г., над этим стоит серьезно задуматься управленческим структурам. Возможно, мы столкнулись с феноменом абсолютной неэффективности и социальной безответственности российских СМИ, которые своей деятельностью не только не способны осуществлять профилактику слухов, вести разъяснительную работу, удовлетворять информационные потребности людей, но и сами сеяли фобии, страхи, тревожные ожидания относительно финансового кризиса.

Обращает на себя внимание тот факт, что в каждом субъекте мишенью слухов были различные банки, всего, по нашим сведениям, напрямую пострадало более 20 банков (т.е. у них произошло массовое, паническое изъятие размещенных населением денежных средств). Таким образом, если правомерно предположение о действиях агрессивных коммуникаторов, они должны были, как по специальному сигналу, начинать генерировать искусственные слухи — каждый в своем регионе России, да еще в отношении различных, ничем не связанных между собой банков. Напрашивается вывод: если мы столкнулись с деятельностью создателей искусственных слухов, их целью был не какой-то конкретный банк, а, по-видимому, вся банковская система России. Этим можно объяснить, с одной стороны, синхронизацию действий по времени, а с другой — отсутствие какой-то одной явной мишени среди банковского сообщества.

В конце 2008 – начале 2009 гг. экономистами констатировалось мощное давление на рубль, в результате чего он девальвировался по отношению к бивалютной корзине («евро – доллар»). Не последнюю роль здесь сыграли тревожные слухи среди населения, заставлявшие людей скупать валюту в обменных пунктах, избавляясь от рублей. В массовом сознании господствовало представление о том, что рубль будет все больше обесцениваться, а значит, необходимо любые денежные накопления срочно переводить в иностранную валюту. В итоге Центробанк вынужден был терять миллиарды долларов и евро в месяц, поддерживая курс рубля и не допуская его резкого падения. Таким образом, налицо уже вполне реальные, а не виртуальные финансовые, репутационные потери от тревожных разговоров, пересудов, домыслов.

¹ Куликов Е.М., Кубякин Е.О. Слухи как коммуникативный и социокультурный феномен современного общества. Краснодар, 2009. С. 100–115.

В ситуации функционирования слухов о мнимом банкротстве различных банков в субъектах РФ сложно выделять только экономические, политические или социальные последствия. Как представляется, эти последствия тесным образом взаимосвязаны в нынешней российской действительности. Экономический кризис и тревожные ожидания — это сегодняшнее проявление массового сознания, его доминантная форма. Многие реальные элементы кризиса имеют в своей основе виртуальные, коммуникативные конструкторы, основанные на панических, упаднических настроениях, опирающихся чаще на домыслы и досужие рассуждения, чем на реальные, объективные факты. Вместе с тем «кризис в сознании» и коллективные тревожные ожидания нашли выражение во вполне реальных экономических факторах. Очевидно, что в слухах заложен значительный ресурс воздействия на сознание и поведение людей. Особенно опасны негативно окрашенные слухи, сеющие панику, фобии. Они способны парализовать работу банковской системы, обвалить фондовый рынок, создать проблемы государственному учреждению или частной компании.

Отметим, что руководству банков стоило больших усилий не допустить реального банкротства вследствие того, что вкладчики, под влиянием тревожных слухов, изъяли из оборота банков огромные денежные средства в течение октября 2008 г. Подобные события вскрывают особенности управления репутацией организаций, компаний в новых социокоммуникативных условиях. PR, безусловно, испытали значительное воздействие изменений, происходящих в обществе в последние 10–15 лет. Возникновение и развитие новых средств коммуникаций видоизменяют характер и способ деятельности специалистов PR. Усиливается роль и таких новых тенденций, как глобализация и современные технологии. Легче стал процесс общения с миром, но сложнее быть услышанным вследствие перегруженности каналов вертикальной коммуникации. Век он-лайн коммуникаций и усиление конкуренции приводит к жесткой борьбе, обуславливающей доступ к СМК. Это объясняется возникновением и развитием новых технологий, обеспечивающих моментальную коммуникацию. События переносятся с молниеносной быстротой, поэтому специалисты по PR должны не только обладать высоким профессионализмом, но и хорошо знать новейшие технологии¹.

Кроме того, возрастает значение точной оценки результатов деятельности. Специалисты по PR, соответствующие отделы, подразделения призваны приносить реальную пользу компании, которую можно было бы объективно зафиксировать. В этой связи активно применяются прикладные социологические исследования, позволяющие определить эффективность PR-мероприятий. Регулярно проводимые прикладные исследования дают возможность оценивать динамику показателей и принимать решение о конечной оценке деятельности PR-подразделений и конкретных сотрудников, отвечающих за те или иные направления коммуникативной деятельности. Управление коммуникациями организации, компании должно опираться на данные прикладных социологических исследований, призванных установить «обратную связь» с целевой аудиторией, получать оперативные данные об угрозах корпоративной репутации.

И, конечно, эффективность PR напрямую связана с умением преодолевать негативное воздействие слухов. Общество XXI отнюдь не лишено этого «информационного вируса», преследовавшего человечество тысячи лет. Напротив, в современном обществе слухи успешно адаптировались к социокоммуникативным реалиям и продолжают, как и раньше, оказывать воздействие на людей, теперь все чаще начиная распространяться в сети Интернет. Это необходимо учитывать специалистам по PR.

Проблема эффективных PR-мероприятий в ситуации функционирования слухов актуальна и для государственных управленческих структур, администраций городов, районов, субъектов РФ. Это еще раз доказали события сентября 2010 г., произошедшие в Краснодарском крае накануне празднования Дня города в Краснодаре, Армавире. Празднования Дня города — важный элемент PR-мероприятий государственных управленческих структур. Однако на этот раз, в дело вмешался «информационный вирус», т.е. слухи.

¹ Почепцов Г.Г. Паблик рилейшнз, или Как успешно управлять общественным мнением. М., 2004.

Известно, что слухи о готовящихся терактах начали распространяться в Армавире, одном из крупнейших городов Краснодарского края в середине сентября 2010 г. Среди распространяемых слухов была история о том, что в городском парке Армавира якобы был найден гроб с гвоздиками и запиской о том, что число жертв в День города будет равным числу цветов. Кроме того, говорилось и о различных надписях на стенах домов, также носящих угрожающий характер. Представители городского ГУВД информацию о возможном теракте и подозрительных находках не подтвердили. Затем волна паники перекинулась на Краснодар, причем слова о «гробе с гвоздиками» повторялись почти дословно¹. Стало очевидным, что затем произошла модификация содержания слуха. В результате в Краснодаре в преддверии празднования Дня города 25–26 сентября 2010 г. стали ходить слухи о том, что якобы возле администрации кто-то видел гроб, набитый уже гвоздями, а не гвоздиками. К гробу якобы прилагалась записка с сообщением о том, что в День города среди горожан будет столько же жертв, сколько гвоздей в этом гробу².

Массовое распространение слухов об угрозе терактов спровоцировало панику среди части населения. Наиболее впечатлительные стали вести затворнический образ жизни, отказались от участия в праздничных мероприятиях, впали в депрессию. Естественно, правоохранительные органы и административные структуры были вынуждены работать в чрезвычайном режиме, учитывая состояние многих людей. Ни о каком позитивном эффекте весьма затратных PR-мероприятий речь уже не шла. День города был безнадежно испорчен и превратился в «головную боль» государственных управленческих структур.

В этой ситуации уместен вопрос: каковы возможности социологии при решении данной проблемы? Например, в арсенале эмпирической и прикладной социологии имеются социально-инженерные подходы, позволяющие диагностировать суть проблемы и выработать рекомендации по недопущению повторения печальных событий сентября 2010 г.

В основе социально-инженерного подхода, применительно к ситуации массового распространения тревожных слухов, может быть положен метод контент-анализа. Это один из способов получения информации об угрозах репутации государственных организаций и частных компаний, исходящих от недостоверной, провокационной информации в глобальной сети Интернет. В «Рабочей книге социолога» контент-анализом называется техника выведения заключения, производимого благодаря объективному и систематическому выявлению соответствующих задачам исследования характеристик текста. При этом подразумевается, что применение данной техники включает некоторые стандартные процедуры, часто предполагающие измерение. Полученные данные должны обладать заданным в исследовании уровнем обобщенности³. К контент-анализу обычно прибегают в следующих случаях:

- а) имеется большое количество несистематизированного материала, причем его настолько много, что становится оправданным применение статистического анализа;
- б) нужна высокая степень объективности анализа информации;
- в) требуется высокая степень точности при сопоставлении однопорядковых данных;
- г) сравниваемые характеристики анализа сопоставимы (например, содержание газетных сообщений – численность подписчиков – их мнение об этих материалах)⁴.

Нами в октябре–ноябре 2010 г. было проведено методически ориентированное исследование, направленное на апробацию методологических, методико-процедурных и организационно-технических моментов ассистирующей деятельности социологов (с применением метода контент-анализа) во взаимодействии со специалистами по PR. В дальнейшем представляется возможным организация и проведение уже собственно социально-инженерного исследования по мониторингу и противодействию распространению слухов в Интернете. Далее приводятся некоторые результаты методически ориентированного исследования и вырабатываются рекомендации специалистам коммуникационного менеджмента.

¹ <http://www.yuga.ru/news/202456/> 25.09.2010.

² <http://www.kuban.aif.ru/crime/news/30951> 25.09.2010.

³ Рабочая книга социолога / Отв. ред. Г.В. Осипов и др. М., 1983. С. 293.

⁴ Зборовский Г.Е., Шуклина Е.А. Прикладная социология: Учеб. пособие. М., 2006. С. 113–114.

В результате проведенного контент-анализа удалось локализовать все основные формы существования слуха об угрозе терактов в Краснодарском крае (таблица 1). Оказалось, что слух функционировал в Интернете в 10 основных формах, самой популярной из которых оказалась «сколько гвоздей в гробе, столько будет и жертв на день города»/«сколько гвоздик в гробе, столько будет и жертв на день города» (29 %).

Таблица 1

Частота упоминаний в Интернете тематических блоков, связанных с функционированием слухов об угрозе терактов при проведении культурно-массовых мероприятий (поисковая система «Яндекс», октябрь 2010 г.), %.

Тематический блок	Частота упоминаний
1. Слух о гробах, слухи о гробах, слухи о гробах в Краснодаре, Армавире, Тимашевске, Курганинске, Темрюке	8,7
2. Гвоздики в гробах, гробы с гвоздиками, фото гроба с гвоздиками	7,3
3. Гвозди в гробах, гробы с гвоздями, фото гроба с гвоздями	6,5
4. Теракт на остановке в Краснодаре, теракт в мкрн. Гидростроителей ¹	2,0
5. Угрозы, сообщения о терактах в Краснодаре, Армавире, Тимашевске, Курганинске	9,7
6. Свидетельства, сообщения о теракте «очевидцев» в Краснодаре, Армавире, Тимашевске, Курганинске ²	14,4
7. Фото взрыва и последствий взрыва в Краснодаре, Армавире и т.д., фото жертв взрыва в Краснодаре, Армавире и т.д. ³	3,6
8. Паника, страх в Краснодаре, Армавире, Тимашевске, Курганинске и др.	7,4
9. «Сколько гвоздей в гробе, столько будет и жертв на день города»; «сколько гвоздик в гробе, столько будет и жертв на день города»	29,2
10. Найден гроб в Краснодаре на Театральной площади, на Красной площади, найден гроб Армавире и т.д.	11,2

На втором месте по степени распространенности свидетельства мнимых очевидцев в разных лексических и фразеологических вариациях (14 %). Также 11 % слухового поля занимали мнимые новостные сообщения о факте обнаружения гроба с угрожающим террористическим посланием. Таким образом, можно восстановить процесс распространения слухов в сентябре 2010 г.: мушкетирование «зловещей», интригующей фразы, получение мнимых «доказательств» от «очевидцев» и их якобы подтверждение в псевдоновостных сообщениях. Именно так оказывалось психологическое давление на пользователей сети Интернет и конструировалась иллюзорная картина террористической активности. Если бы эти данные оказались в распоряжении специалистов коммуникационного менеджмента в момент зарождения и начальной фазы распространения слухов, у них были бы реальные возможности, приняв ряд контрмер, не допустить паники среди населения.

Нам удалось воспроизвести также и хронологию интенсивности распространения панических слухов об угрозе терактов в Краснодарском крае в сентябре 2010 г. В качестве примера мы взяли наиболее распространенную формулировку слуха «Сколько гвоздей в гробе, столько будет жертв на День города» (таблица 2). Конечно, стоит сразу оговориться: часть сообщений были удалены пользователями и оказались недоступны для нашего анализа, проведенного в октябре-ноябре. Также нужно указать, что у исследователей нет соответствующих полномочий для просмотра архивов сайтов и восстановления удаленных сообщений, фото- и видеоматериалов.

¹ Злоумышленники просто скопировали сообщения о реальном теракте в Краснодаре в 2003 г., выдав старую информацию за сенсационную новость 2010 г.

² Это уже слухи, в действительности ничего подобного в 2010 г. не произошло.

³ Аналогично размещенные фотографии никак не связаны с событиями в Краснодарском крае.

Частота упоминаний в Интернете наиболее распространенной формулировки слуха: «Сколько гвоздей в гробе, столько будет жертв на День города»
(поисковая система «Яндекс», октябрь 2010 г.)

Дата	Число упоминаний фразы
15.09.2010	137
16.09.2010	119
17.09.2010	125
18.09.2010	214
19.09.2010	231
20.09.2010	192
21.09.2010	180
22.09.2010	228
23.09.2010	237
24.09.2010	288
25.09.2010	372

Вместе с тем, даже имеющиеся в открытом доступе данные позволяют обнаружить точку зарождения слухов — 15 сентября 2010 г., когда в течение одного дня 137 раз муссировалось угрожающее послание. «Стимулом» для активности распространителей слухов (по нашим данным) стала фотография гроба, выложенная на одном из сайтов, базирующегося далеко за пределами Краснодарского края. Естественно, данная фотография никак не была связана ни с Армавиром, ни с Краснодаром, т.е. вымышленные события обрастали якобы «доказательствами»¹. В результате угрожающая формулировка: «Сколько гвоздей в гробе, столько будет жертв на День города» нашла «документальное» подтверждение, в представлениях пользователей Интернета, которые стали активно обмениваться непроверенной провокационной информацией. Следующий период времени: 16–17 сентября отмечен падением интереса интернет-пользователей к данной теме. Затем произошел «всплеск» активности коммуникаторов (18 сентября 214 сообщений, 19 сентября 231 сообщение). По-видимому, в этот период времени проявилась активность лидеров мнения, за несколько дней переосмысливших информацию и активно включившихся в ее обсуждение, либо злоумышленники², организовавшие данную провокацию, предприняли попытку искусственного повышения интереса к данной теме³.

Однако наибольшая активность коммуникаторов пришлась на период 22–25 сентября 2010 г. Именно в это время слух о гробах с угрожающими посланиями «перекочевал» в Краснодар — столицу Краснодарского края. Безусловно, эта новость охватила значительную по численности аудиторию, слух достиг своего пика. Дальнейшее затухание слуха объясняется просто: День города прошел, теракт к счастью не случился, интерес к данной теме благополучно иссяк, и в дело «развенчания» слухов включились, правда, со значительным опозданием, городские и краевые СМИ⁴, люди убедились в недостоверном характере информации. Однако негативное воздействие на репутацию государственных управленческих структур слух оказал, и именно это призван не допускать или сводить к минимуму коммуникационный менеджмент.

¹ Фотография позже была кем-то удалена, ее «следы» мы обнаружили только в полемике собеседников, ссылавшихся на фото гроба, якобы найденного в Армавире (а затем в Краснодаре).

² Анализ событий 2010 г. в разных регионах РФ показал, что аналогичный слух распространялся в Элисте, Астрахани, Таганроге и др. городах. Кто-то, зная расписание «Дней города», просто актуализировал слух применительно к тому городу, в котором скоро ожидалось торжественные мероприятия.

³ Они могли, например, по предварительному сговору, начинать обмениваться комментариями на тему «гробов с угрожающими посланиями», втягивая в дискуссию ничего не подозревающих пользователей Интернета, у последних при этом возникала иллюзия естественного хода обсуждения.

⁴ Часть из них, например, газета «Краснодарские известия» со ссылкой на краевое УВД поспешили сообщить: «Распространитель слухов найден!». Под подозрение попал один из сотрудников администрации Краснодара, отправивший слух в виде более 100 SMS-сообщений родственникам, друзьям, знакомым. Однако если бы государственные управленческие структуры и СМИ сотрудничали с исследователями-социологами, они наверняка бы знали о реальной картине происшествий и не делали поспешных заявлений.

Конечно, традиционный контент-анализ, к сожалению, уже не соответствует реалиям XXI в. Перспективным выглядит применение компьютерного контент-анализа, учитывая огромные массивы информации, подлежащей анализу и специфику глобальной сети Интернет. Целью прикладного социологического исследования может стать как восстановление хронологического порядка распространения негативно окрашенных, панических, провокационных слухов, ухудшающих корпоративную репутацию (поиск Интернет-ресурсов, где произошел генезис слухов, составление базы данных наиболее активных интернет-пользователей, которые «внесли вклад» в процесс распространения слухов), так и режим мониторинга Интернета посредством интеграции программы контент-анализа в основные поисковые системы.

О системе мониторинга интернет-пространства стоит сказать подробнее. Имеющиеся в распоряжении тех же российских банков компьютерные программы поиска информации по ключевым словам (обычно по названию банка), выглядят далекими от совершенства и в настоящее время поставляют коммуникационному менеджменту «горы информационного мусора», который практически никогда не подвергается детальному и осмысленному прочтению. Компьютерный контент-анализ призван в режиме реального времени классифицировать информационный поток, так или иначе связанный с компанией, на три категории: а) оказывающий позитивное воздействие на репутацию; б) не оказывающий ни позитивного, ни негативного воздействия; в) оказывающий негативное воздействие. При этом программа должна быть интегрирована в поисковые системы, отвечающие за неформальную коммуникацию (чаты, блоги, форумы и т.д.), поскольку именно здесь часто возникают слухи, угрожающие корпоративной репутации.

Очевидно, что в случае успешного осуществления мониторинга интернет-пространства с помощью метода компьютерного контент-анализа, становится возможным не только оперативно реагировать на изменение ситуации в коммуникационных системах, но и повышать эффективность управленческих PR-мероприятий.

Статус и возможности слабоструктурированного семантического пространства текста

Леонтьева Нина Николаевна,
*Московский государственный университет
им. М.В. Ломоносова*

Философия системы, понимающей текст

В лингвистике менее всего поддается формализации *семантический уровень*, тогда как именно семантика отвечает за автоматизацию понимания и она более всего востребована интеллектуальными системами разного рода. Это было понятно и 50 лет назад, когда Вяч. Всев. Ивановым была поставлена задача: «искать элементарные смыслы». Современная лингвистика имеет множество наработок в грамматике и в лексикологии, и в тонком анализе смыслов, передаваемых отдельным предложением, и в семантике каждого отдельного слова и т.д. Но, несмотря на локальные и часто эффектные отдельные успехи, сфокусировать эти знания в полной лингвистической модели понимания и ее компьютерной реализации она пока не смогла. Классическая лингвистика не снабдила нас необходимой теорией семантики и формальным аппаратом, нужными для решения современных задач. Это объясняется и эффективной сложностью комплексной задачи автоматического понимания *естественных текстов (ЕТ)*.

Здесь нужна дисциплина, умеющая работать на стыке наук. Ведь семантика не ограничивается связями только с предшествующими ей уровнями анализа текста, но должна учитывать и внетекстовые знания, без которых *автоматическое понимание текстов (АПТ)* не может быть полноценным. Именно отсутствие принципиальных проработок всех таких связей вынуждает прикладного лингвиста вновь и вновь обращаться к проблемам семантического анализа целого текста (или ЕТ). Самой большой проблемой, с которой прикладная лингвистика не может справиться, является *механизм взаимодействия текста и внешней среды*, к

каковой мы отнесем разные специальные и энциклопедические знания, как общие, так и индивидуальные: это и взгляды, убеждения, оценки и даже эмоции человека или иной интеллектуальной системы, «понимающей» тексты. Без включения такого компонента в Систему трудно говорить о Смысле. Назовем его *воспринимающим устройством (ВУ)*.

Когда человек извлекает из текста «свою» информацию, или находит *свой смысл*, он может распорядиться им по-своему. Он может использовать новое знание в своих интересах и тем самым выпадает из компьютерной системы, либо «пускает в жизнь» эту информацию. Второй способ, в отличие от первого, индивидуалистического, назовем «общественным поведением». Система АПТ должна обладать именно *общественным поведением*, предоставляя возможность пользоваться знаниями, добытыми из текстов, каждому индивидууму и всему социуму.

Главное условие, которое делает создание модели компьютерного понимания ЕТ *реалистичным*, состоит в том, что все три названные компонента — исходный текст, привлекаемые знания и получаемая информация — должны быть *вербальными*, т. е. приниматься к рассмотрению только в виде текстов, хоть и понимаемых достаточно широко (таблицу или структуру тоже будем считать «текстом»). Индивидуальные и специальные знания и даже реакции ВУ на результат понимания также должны удовлетворять условию *текстовости*. Для взаимодействия всех компонентов необходимо, чтобы все они были представлены на некотором *метаязыке*, едином для всех участников работы Системы. Только тогда можно говорить о некоей алгебре текстов и опираться на важные свойства текста (связность, наличие фокуса и другие). Проследим те процедуры АПТ и особенности словарной информации, которые предусмотрены и были частично реализованы в нашей *информационно-лингвистической модели (ИЛМ)*.

Самым малым элементом модели будем считать слова естественного языка (ЕЯ). Ну, а если слова, то опять лингвистика, и больше всего *Словарь*. «Элементарные смыслы» тоже используются нами, но лишь как технический прием при различении значений слов в словаре (слово в определенном значении называется Лексемой и составляет вход в семантический словарь) и при описании смысловых предсказаний для каждой валентности единиц словаря. Работа с крупными компонентами тоже имеет лингвистический характер из-за установки на вербальный характер всех структур ИЛМ.

Итак, философия системы — это архитектура системы АПТ. В ней важны содержание каждого компонента и цель каждого процесса, в том числе трудоемкость ее достижения. Пока мы можем говорить лишь о моделях и архитектуре комплексов, описывающих автоматическое или механическое «понимание» текста. Но сама модель системы АПТ может влиять на уточнение теории семантики.

Контуры информационно-лингвистической модели (ИЛМ)

Наибольшие трудности АПТ связаны с уровнем семантического анализа (СемАн), который ответственен за получение Смысла (из Текста). Действительно, лингвисты имеют дело с текстами и смыслами. Позволю себе сравнить две модели.

Имя главной отечественной модели и теории СМЫСЛ $\Leftrightarrow$ ТЕКСТ (МСТ, ТСТ) всеобъемлюще. Однако в разных прикладных задачах обе части (Текст и Смысл) уточняются по-разному. Собственно МСТ моделирует автоматическую генерацию текста из некоего семантического представления (СемП), в котором сам смысл постулируется, т.е. СемП является скорее представлением ученых (в частности И.А. Мельчука как автора МСТ) о том, как устроен Смысл целого текста. Видимо, именно к такому СемП надо стремиться и при анализе текста. Наиболее грамотный анализ в соответствии с ТСТ выполняется в рамках создания систем машинного перевода и других интеллектуальных систем коллективом под руководством академика Ю.Д. Апресяна. Реально результатом анализа являются структуры, называемые автором семантическими (СемС), но я называю их синтактико-семантическими представлениями (СинСемП), так как в основе остаются обогащенные семантикой синтаксические представления предложений. Результатом анализа целого текста будем поэтому считать последовательность СинСемП всех предложений текста. Взятые вместе, Анализ текста по

Апресяну плюс Генерация текста по Мельчуку образуют строго *лингвистическую* модель понимания текста. Добавим, что в конкретных системах реализации МСТ анализ и синтез опираются на Толково-комбинаторные словари (ТКС) нового типа.

Наша конкретная установка — автоматическое понимание текста — моделирует, в отличие от МСТ и ТСТ, путь от реальных ЕТ к извлекаемой из них Информации, отсюда и название «информационно-лингвистическая модель (ИЛМ)». Поэтому нам так важно разделить три понятия: Содержание текста, Информация и Смысл. Понятие СемП, называемое авторами МСТ Смыслом текста, я считаю полным и подробным, со всеми лингвистическими нюансами, пофразным *Содержанием* текста. Структура целого текста у нас отличается от СемП в МСТ: это не окончательная, а промежуточная и реально достижимая структура, правильнее назвать ее *семантическим пространством текста*. И Смысл мы определяем иначе — как вторичное по отношению к понятию Информация. Вычисление Информации требует активного учета воспринимающего устройства (ВУ), будь то конкретный пользователь или интеллектуальная Система, для которых данный текст *имеет Смысл* только потому, что содержит *Информацию*, важную для ВУ.

Это ВУ привлекается в систему понимания текста не в своем физическом виде, а в виде текстов, в которых сформулированы информационные интересы, вопросы или просто включен эталонный для ВУ фрагмент текста; все они названы «*встречными текстами*», так как мы имеем дело с вербальной моделью. Если к семантическому пространству текста добавить те встречные тексты, которые сформулировал или выбрал конкретный пользователь, мы получаем *информационное пространство текста*.

Разница между содержанием исходного текста и содержанием встречного текста даст нам ту величину (структуру), которую можно назвать Информацией из данного текста (Т1) для данного ВУ. Информацию, или индивидуальное Информационное представление (ИнфП) можно вербализовать, построив текст Т2, отличный от Т1. Тем самым в составе ИЛМ даже в пределах одного естественного языка (ЕЯ) различаются Т1 и Т2, т.е. результат понимания Текста1 приводит к Тексту2, отличному от исходного Текста1:

(Текст1)

СинП предложений – СемП высказ-й – СемГраф текста

СинСемП – Сем(Ситуативное)П – СемПространство текста

**Расширенный («встречными текстами») СемГраф
(=Информационное Пространство текста)**

**ИнфП – Информация – СМЫСЛ – (Текст2)
(для данного читателя из данного текста)**

Информация — хрупкая структура: после того как она произвела изменения в Базе знаний Системы или ВУ, она перестает быть Информацией; при другой установке или для другого пользователя может быть извлечена из того же текста другая Информация (и построен другой Текст2). ИнфП можно добавить к имеющимся у ВУ Знаниям (они тоже присутствуют в системе вербально: в виде текстов, таблиц, семантических сетей и т.п.), и тогда в следующий раз он тоже получит иную, новую Информацию, другой Смысл и новое оформление его в текст Т2. Это определение Информации по Шрейдеру. Он определял ее как разницу в Тезаурусах текста и читателя, причем Тезаурусом он называл запас Знаний, имеющихся у объекта понимания (ЕТ) и у читателя (ВУ), соответственно. Такая «алгебраическая» процедура вычитания «текстов», конечно, нуждается в едином *метаязыке* для всех структур семантического уровня и их единиц.

Итак, в нашем определении Смысл текста НЕ совпадает с его СемП, которое и объявляется Смыслом в современном классическом подходе (ТСТ). Тем не менее термин СемП мы будем считать родовым понятием для всех смысловых структур, включая и наше промежуточ-

ное Семантическое пространство, и те, которые отображают только главную или заказанную часть смысла, как в ИПС и других системах «извлечения знаний из текста» (Information Extraction systems).

Уточним и понятие Текст в нашей модели. Одно предложение, да еще типа *Джон любит Мэри* — это еще не текст, а псевдотекст, по В.А. Звегинцеву. Чтобы их различить, мы используем термин ЕТ и останавливаемся на разборе наиболее трудных примеров, игнорируемых лингвистами как неправильные, неполные и т.п. Говоря о компьютерной системе понимания, мы имеем в виду непрепарированный корпус, или массив, ЕТ: только для них и нужна Система АПТ и модель, по которой ее можно строить. Система должна принимать на входе реальные Тексты (=ЕТ) со всеми их недостатками и анализировать их со всеми присущими текстам атрибутами, учитывая семантику самого массива, наличие заголовка, имен авторов, членимость на более мелкие композиционные и содержательные единицы и прочее. (Даже фразы-реплики типа «*Моя твоя не понимай*» семантика должна интерпретировать, хотя Синтаксис вправе отказаться от создания СинП для него).

Итак, при установке на общественное поведение Системы АПТ и социальную заинтересованность в результатах ее работы, будем рассматривать ТЕКСТ в его естественном виде, а Информацию понимать как новое знание — только оно и представляет СМЫСЛ для пользователя. Эти установки ИЛМ позволяют сформулировать требования «сверху» к единицам лингвистического анализа: они должны быть СРАВНИМЫ с единицами других текстов и со структурами представления знаний. Ясно, что в своем первозданном и «непричесанном» виде разные ЕТ не допускают формального сравнения. Вот почему так важно предложить единый «усредненный» *метаязык*, на который можно было бы переводить разные (по форме и содержанию) источники знаний.

О семантическом пространстве текста как словарном отображении

Первичный семантический анализ, который является основным этапом построения *семантического пространства текста*, ведется для каждого выделенного сегмента СинП или СинСемП методом сравнения всех слов с лексемами семантического Словаря. Каждая лексема имеет в Словаре *семантические характеристики* и *смысловые валентности* (СХ, ВАЛ), кроме сведений, необходимых для дальнейших операций. Валентности стремятся перейти в конечном счете в имена дуг, или семантических отношений, между узлами строящегося СемП текста:

$$\begin{array}{c} \text{СемО} \\ \text{СемУ_А} < \text{-----} \text{СемУ_В.} \end{array}$$

Здесь СемУ_А и СемУ_В — семантические узлы (СемУ) с лексическими ядрами А и В соответственно; СемО — связывающее их двуместное отношение. Это же можно записать такой формулой: СемО(СемУ_А, СемУ_В), или Р(А, В); читается «А связано с В отношением Р», где В — семантически главный член по умолчанию.

Одна из задач анализа — преобразовать лексемы первичного СемП в сложные СемУзлы, описывающие Объекты со всеми присущими им атрибутами, или даже целые Ситуации. Простые предложения желательно свернуть в Ситуативные Представления (СитП), по возможности определив СемО и между ними. В идеале множество СитП, образующих иерархию, перейдет в окончательный СемГраф, составив полноценное *сжатое Содержание* текста. Эти задачи определяют и состав сведений, записываемых в Словаре при каждой лексеме. Информация в нем устроена так, чтобы обеспечить именно смысловое сжатие текста.

Валентность в нашем подходе — прежде всего единица семантической природы. Для каждой входной лексемы (С), имеющей определенное значение, она задает возможные семантические связи с другими узлами (обозначим их А1, А2...Ак), например, АГЕНТ(А1, С); ПРИЧИНА(А2, С); АДРЕСАТ(А3, С); ВРЕМЯ(А4, С); ЦЕЛЬ(А5, С). Валентности предсказывают, по каким связям (или СемО) должно идти в норму развитие текста, в этом их отличие от СинСем валентностей, действующих в пределах фразы.

Множество значений словарного поля «ВАЛ»ентности в виде предсказываемых смысловых связей найденного слова с другими словами (включая способы реализации в тексте) оказывается основой искомого метаязыка. Средствами метаязыка СемО можно описывать *любые связи в тексте*, а не только реализующиеся в пределах предложения. Такой семантический метаязык можно считать *информационным языком-посредником (ИЯП)* системы АПТ. Информационным он назван потому, что формула $P(A,B)$ является *минимальной единицей передачи информации*, а именно задачу вычисления и передачи Информации (уже в терминологическом смысле) призвана решать Система. Как язык ИЯП имеет свою Грамматику (синтагматику и парадигматику). А посредником он считается по своей главной функции – обеспечивать обмен данными между всеми компонентами Системы. Переводом всех элементов текста на принятый ИЯП (при этом мы опускаем многие сложности, а также и другие источники дополнения списка СемО) можно получить ту структуру, которая и образует семантическое пространство (СемПрост) текста.

Часть объявленных в Словаре валентных связей, заполненных при анализе словами предложения, переходят в структуру СемП, незаполнившиеся тоже переносятся, но с признаками неполноты. Эти элементы общей структуры текста, т.е. имена связей, и с содержательной точки зрения не менее информативны, чем узлы. Неполные формулы типа $P(A,?)$ или $P(?,B)$ или $P?(A,B)$ образуют область Незнания и требуют заполнения в ходе анализа уже целого текста. *Неполнота* оказывается НЕ недостатком, а конструктивным фактором анализа. Устранение неполноты, а также избыточности часто достигается единой процедурой сложения двух неполных формул из разных фраз, при этом получаем одну полную по правилу: $P(A,?) + P(?,B) = P(A,B)$. Проиллюстрируем это правило на простом примере: *Он все-таки послал письмо. Но только брату. Матери он не решился послать.*

АГЕНТ (ОН,послать) --- Сит1

РЕФерент (? , ОН)

ОБЪЕКТ (письмо, послать)

СОДЕРЖ (? , письмо)

АДР (? ,послать)

ЦЕЛЬ (? , послать)

АДР (брат,?) --- Сит2

КВАНТР (только, брат)

АГЕНТ (ОН,послать) --- Сит3

РЕФ (? , ОН)

3. АДР (мать,послать)

4.МОДальность (не_решаться,послать)

5.ПРИЧина (? , Сит3)

Здесь ВАЛ=АДРесат, оставшаяся неполной в Сит1, заполняется лексемами из Сит2: АДР (? ,послать) + АДР (брат,?) = АДРесат (брат,послать). Другие неполные формулы порождают вопросы: Кто Он? О чем письмо? Почему не послал матери? и т.п. Если в результате анализа всего текста некоторые связи остались незаполненными или ненасыщенными, они элиминируются из СемП. В дальнейшем такие данные результатов анализа могут использоваться для коррекции поля валентностей в Словаре.

Итак, устранение неполных формул приводит к сжатию содержания. Чтобы потери информации при сжатии были контролируемы, т.е. не отбрасывалось то, что относится к основному содержанию текста, нужно выработать критерии сжатия. Это эквивалентно определению *степени информативности* строящихся текстовых единиц.

Критерии информативности единиц, определяющие сжатие, задаются частично самой классификацией лексем в Словаре: лексемы, переходящие в СемУзлы типа Ситуация и Объект в качестве лексического ядра, имеют по умолчанию максимальный ВЕС (5); лексемы-признаки и абстрактные действия имеют ВЕС 4, и т.д. Минимальный вес приписан служебным элементам. Лексемы, образующие имя отношения в СемГрафе текста, тоже ранжированы. Так, максимальный вес в словаре имеют единицы, входящие во фрейм СИТуации (Вре-

мя, Место, Причина, Цель); но ВЕСа отношений могут быть повышены в ходе анализа, если их аргументами оказываются единицы с высоким весом.

Итак, первичное ранжирование единиц по степени информативности дает Словарь, а носительный информационный вес узлов и связей, а также составленных из них единиц может «вычисляться» и далее уточняться при семантическом анализе текста. При этом анализ обращается и к правилам вывода, записанным в терминах СИТуаций в том же Словаре; в нем имеется также много дополнительных предсказаний о том, как может развиваться Ситуация в тексте, какие связи могут ей предшествовать и следовать, и т.д. Важно, что все такие данные записаны на одном метаязыке и могут быть непосредственно перенесены в СемП. В настоящее время Русский Общесемантический Словарь – РОСС, впоследствии переименованный в РУСЛАН (Русский Словарь для Анализа), законсервирован ввиду отсутствия экспериментальной базы. Он содержит больше 12 тыс. единиц (однословные лексемы описаны в составе Базы данных Росс, а Словари отношений существуют в текстовом виде, но их описания сделаны в таком же формате).

Ситуативное представление высказывания и текста

Итак, первичный семантический анализ текста опирается на Словарь, применяя его к синтаксическим структурам — СинП — для всех выделенных предложений или их фрагментов. Наиболее удобным входом в СемАн представляется такой результат СинАн, который сегментирует текст на простые предложения и строит СинП для всех простых предложений текста (см. доклад Т.Ю. Кобзаревой в материалах Конференции). Уже после вычленения простых предложений в составе сложных и сборки разорванных частей в целое предложение мы получаем в общем случае сложный граф, где придаточные могут зависеть не от всего включающего их предложения, а от отдельных его слов. Этим уже намечен состав сложного СемУзла. Каждое простое предложение интерпретируется в терминах семантических единиц, главной из которых является лингвистическая Ситуация, вводимая нетерминальным символом СИТ. Для каждого отдельного фрагмента СинП происходит сборка лингвистических единиц типа СИТ сравнением с заданной типовой структурой простого предложения (или с фреймом Ситуации).

Этапы преобразований, на которых происходит сборка единиц типа СИТ, можно назвать ситуативным анализом (СитАн); он собирает те кирпичики, из которых строится знание. Сит — та «средняя» единица, которая еще сохраняет черты единиц поверхностного уровня. Первичная семантическая интерпретация проводится по всем фрагментам СинАн, в том числе и отрезкам, не включенным ни в одно СинП. Каждому СемУзлу, оставшемуся изолированным, присваивается валентность на вхождение в какую-то СИТ (все они нумеруются при анализе).

В полном виде узел СИТ имеет ядро, подчиненное нетерминальному символу СИТ. Чаще всего это лексическое ядро (ЛЯ) — в стандартном случае ЛЯ Ситуации образуется из сказуемого простого предложения (или простейшего Высказывания), становясь многоместным предикатом с СемУзлами, присоединяемыми по его ситуативным связям. (Только эти СемУзлы мы соглашаемся называть в нашей модели актантами СИТ). Многие структуры Сит при этом остаются локально (или временно) неполными.

Для символа СИТ обязательно указание модальных и идентифицирующих отношений: МОДальность (?, СИТ), ИДЕНТификация (?, СИТ). Ему непосредственно могут быть подчинены семантические узлы, обозначающие Время, Место, а также разные Условия и Обстоятельства, характеризующие Ситуацию в целом. На первом шаге интерпретации ими становятся синтаксические группы, являющиеся обстоятельствами, сирконстантами синтаксической структуры, в дальнейшем некоторые из них могут перейти в ранг актантов лексического ядра, если у него осталась незаполненной близкая по семантике ситуативная валентность.

Упомяну один нестандартный компонент Высказывания, обязательный для описания семантики именно связного текста: это отношение СОЛИСТ(Ак, ВЫСК), оно фиксирует тот СемУзел (Ак), который «подхвачен» следующим за данным высказыванием. Это отношение апостериорное (его можно поставить, только когда проанализировано Высказывание, сле-

дующее за данным), показывающее, по каким темам идет развитие текста. Чередование Солистов в семантических структурах и синтаксическое деление предложения на тему-рему слабо коррелируют друг с другом: ведь реальные тексты часто имеют самые причудливые рисунки связности.

Рассмотрю один нестандартный пример того, как строится СитП (приведу цитату из публикаций автора):

«Ночь. Улица. Фонарь. Аптека.» Эти четыре высказывания вводятся четырьмя единицами СИТ: СИТ1 уточняется семантическим отношением ВРЕМЯ(*ночь*, СИТ1), СИТ2 — отношением ЛОК(*улица*, СИТ2), тем же отношением уточняется СИТ4: ЛОК(*аптека*, СИТ4); каждая из этих семантических формул неполна: во всех Ситуациях отсутствует лексическое ядро (ЛЯ). Третье высказывание будет неполным еще и по имени СемО, связывающего лексему *фонарь* с отсутствующим ЛЯ ситуации: ?СемО(*фонарь*, ЛЯ(СИТ3)): по грамматике символ СИТ не может подчинять семантический узел, обозначающий 'предмет'. Если объявить эти четыре предложения одним высказыванием (а на то есть формальные основания), Система сделает вывод, что речь идет об одной СИТ (т.е. СИТ1 = СИТ2 = СИТ3 = СИТ4), объединит все 4 высказывания в одно и может искать ЛЯ(СИТ) дальше по тексту либо создавать гипотезы о нем. Анализ следующей строки (следующего высказывания) может дальше расширить границы уже построенного высказывания».

Уже начальные работы по семантической интерпретации связей СинП привели нас к выводу о том, что переход от синтаксической структуры предложений к структуре целого текста имеет характер *скачка*: он сопровождается сменой метаязыка, изменением иерархии связей и возможной потерей части информации; но он включает и синтез новой информации. Это хорошо иллюстрируется нашим типовым примером из текста рецепта: *Ложку травы залить двумя стаканами кипящей воды*. Если просто переназвать связи правильного СинП этого простого предложения семантическими именами и после этого сжать полученный СемП по первому правилу сжатия («убирать зависимые члены структуры, оставляя главные»), мы получим реферат «Ложку залить стаканами». СемАн, интерпретируя узлы, построит СемУзлы «Трава (в количестве одна ложка)» *заливается водой* (в количестве два стакана)». При этом *кипящая вода* может быть заменена кратким синонимом *кипяток*, будет восстановлена формула МОДАЛЬНОСТЬ(*необходимо, заливать*) – в соответствии со значением формы Инфинитива; добавится формула ЦЕЛЬ(? , НЕОБХ), в соответствии с валентностью обобщающей лексики НЕОБХОДИМОСТЬ.

1. Колич (ложка, трава)
2. Кон.т (1., заливать)
3. Колич (два, стакан)
4. Колич (3., вода)
5. Призн (кипящий, вода)
6. Модал (НЕОБХ., заливать)
7. Цель (? , заливать)
8. Об (3., заливать)
9. Агент (? , заливать)

Опуская детали, полученный фрагмент СемПрост можно перевести в СитП, в котором лексическим ядром СИТ будет действие *заливать* в модальности *необходимо*.

При всех следующих преобразованиях СемПрост метаязык не меняется: средней единицей в нашей модели является СИТ, минимальной остается формула P(A,B).

При этом СИТ обладает всеми необходимыми свойствами единиц высшего уровня: в структуре СИТ отражена связь с другими высказываниями и всей структурой текста. Если извлечь как цитату или для помещения в базу знаний любой относительно законченный фрагмент структуры, содержащий хотя бы одну единицу СитП, он должен сопровождаться всеми атрибутами текста (минимум названием текста и именем автора).

Множество всех Ситуаций, точнее, последовательность крупных единиц типа СИТ, образует СитП текста. Если вынести из него деление текста на предложения, абзацы, а также дру-

гие сведения о физическом составе текста в отдельную (композиционную) структуру, то мы выходим в пространство ситуаций (СитПрост), представляющее сжатое содержание объекта «целый текст». С единицы СИТ начинается синтез единиц типа Факт.

В итоге анализа целого текста формируется Текстовый Факт (ТФ), куда входит и иерархия СИТ, и отдельные единицы, получившие большой ВЕС при анализе. Строится избыточное СитП и еще более сжатое тематическое представление текста (ТемП как сигнальная информационная структура), и они уже могут выступать как полномочные представители данного текста во внешней среде. Тем самым сохраняется экология текста, позволяющая выделить самые важные (не по статистике, а по содержанию) утверждения текста – по таким сюжетам можно уже давать полноценные ответы.

Только после этого можно сравнивать ТФ, СитП и более мелкие его единицы с подобными структурами внешней среды: структурами предметной области, вопросов к данному тексту, текстовому составу баз данных, если они уже построены в системе.

О синтезе информации

Описанный выше локальный анализ, продолженный глобальным, применяется и ко всем «встречным» текстам, с которыми пользователь приступает к поиску «своей» информации. Структуры (по возможности СитП), будучи добавлены к СемГрафу исходного текста, расширяют его до информационного пространства (ИнфПрост). В нем и должен происходить поиск нужной пользователю информации, этот этап можно уже назвать *информационным синтезом*. Главная единица СИТ встречного текста получает максимальный ВЕС и объявляется СОБЫТИЕМ, в пользу которого должен происходить пересчет весов в этом ИнфПрост.

Динамика построения текстовых структур такова:

СемУ – СемО – ЭСит – Сит – Соб – ТФ – Факт ПрО.

Эту цепочку преобразований нужно понимать так: при построении семантического узла (СемУ) в соответствии со словарным описанием все валентности СемУ (т.е. формулы вида СемО(А,В) становятся элементарными ситуациями (ЭСит). Некоторые комбинации элементарных единиц ЭСит образуют единицу текстового представления СИТ. Некоторые Сит могут быть уточнены как единицы Соб (события), если во встречных текстах ВУ они были выделены как главное содержание того, что его интересует. Некоторые Сит или Соб переходят в единицы, называемые «Текстовыми фактами» (ТФ). Некоторые ТФ могут перейти в базу знаний определенной предметной области (ПрО) как единицы Факт(ПрО), тем самым они образуют фрагмент Грамматики исследуемой предметной области. Если заданная единица СОБ имеет личный характер (например, некоторый пользователь собирает сведения не по главным ядрам СИТ, а по каким-то их характеристикам типа «*начала всех событий*»), он соберет индивидуальную базу данных.

При анализе и подтверждении из других документов, вызывающих доверие, Факты ПрО могут получить статус *ФАКТ* в модальности «истинно». Такой ход анализа и соответствующие выводы (накопление вербальных баз знаний) очень важны при разработке информационных систем для гуманитарных наук.

Генерация новых текстов на основе полученных в сжатом виде текстовых знаний даст реферат исходного текста (Т2). На основе сжатого СемП или его разновидностей (СитП, ИнфП) можно построить хороший реферат (или Т2) текста. Тут мы вправе требовать стерильно правильных предложений и текстов, хотя и здесь можно построить много вариантов (лексических одежд и синтаксических структур). А требовать, чтобы на вход в систему АПТ допускались только полные и правильные предложения, – все равно что в медицинской сфере допускать к врачу только совсем здоровых людей и отказывать всем немощным и больным.

Заключение

Как бы мы ни определяли новую науку — математическая, структурная, прикладная, компьютерная, кибернетическая и т.п., — это прежде всего лингвистика. И если даже чисто лингвистический анализ и только грамотных текстов (здесь я вновь обращаюсь к докладу Т.Ю. Кобзаревой) приводит нас к множественной, неоднозначной и многомерной структуре

СинП, то лучше называть такое СинП синтаксическим Пространством. Тем более этот термин подходит к результату следующих уровней понимания: Сем-, Сит- и Инф-пространство. На самом деле разновидностей семантических структур гораздо больше, как промежуточных, так и окончательных. Так, например, можно построить частные структуры за счет расчленения общего СемПрост на подструктуры по какому-либо признаку, например, по единству темы или по любому выделенному пользователем Событию или по сочетанию признаков – все такие структуры являются Пространствами; к ним применимы дальнейшие процедуры «идеализации», в частности сжатия, устранения противоречий, попытки выявить содержательный центр текста и многие другие.

Методику работы с такими сложными и слабоструктурированными объектами, особенно при их сравнении и в их взаимодействии (а таков именно процесс понимания текста), можно нащупать, а далее уточнять только в реальных экспериментах по АПТ.

СемПрост — это структура, свойства которой отличны от известных моделей, как собственно лингвистических, так и информационных, но частично совмещают их. Уже начиная с СинП, следующие за ним Сем-, Сит- и Инф-пространства — это в норме неточные, неполные, часто неоднозначные структуры. Почему мы продолжаем относиться внимательно к таким «дефектным» структурам?

Во-первых, это первая реальная структура *целого*, которую можно строить и работать с ней, не дожидаясь той заветной поры, когда лингвисты закончат описание всего лексического богатства языка. Понятно, что никогда Словарь системы не будет полным, так как язык развивается, он меняется не только по составу, но и по системе значений лексем, фразем и других единиц. Особенно подвержены изменениям лексика и терминология разных разделов науки, не считая того, что возникают совсем новые направления в науке, технике и социальной среде. Интеллектуальные возможности ИПС и частных профессиональных систем давно исчерпаны: нас уже не могут удовлетворять вопросы в виде неоднозначных ключевых слов и ответы в форме выдачи миллиона текстов, удовлетворяющих формально таким запросам.

Во-вторых, объемы текстовых материалов так возросли, что никакой аналитик не может уже справиться с их обзором. Поэтому самой важной я считаю возможность осуществлять процессы содержательного сжатия текста. Автоматический анализ текста не бывает без потерь, но важно сохранить *главное* в содержании текста, в том числе обязательно и *связи* между наиболее информативными Объектами и Ситуациями. Без развитого лингвистического анализа, общего для всех текстов на данном языке и к тому же учитывающего жанры и специфику разных текстов, невозможно построить сколько-то содержательные структуры. А ждать, пока лингвистическая теория справится со всеми необозримыми пока трудностями полного и непротиворечивого описания всех языковых явлений, специалисты в разных ПрО уже не могут, они начинают придумывать свои «велосипеды», свои лингвистические теории и новые «семантики», мало совместимые друг с другом.

Часто можно читать работы по морфологическому анализу словоформ текста, а в заголовке объявлен «Смысловый анализ текста» или даже «Понимание текста». Здесь взята слишком высокая нота; до понимания еще слишком далеко. Поиск средствами ИПС в Интернете тоже слабо приближается к тому, что можно считать Пониманием текста.

Так, на запрос «магазин свет плафоны розетки Москва» я получила следующий ответ: Перевернутая салатница стала **плафоном**, **розетки** для варенья — подставками для лампочек, пробка графина — подвеской. Огромные окна по всем четырём сторонам дома позволяют дневному **свету** полностью владеть пространством. Естественно, что такое «понимание» не устраивает пользователей; мне нужны были адреса магазина «Свет».

Какие возможности дает подобная, хоть несовершенная и неидеальная, структура Семантического Пространства, если она построена для большого реального текста? И с какой целью мы моделируем работу с такими нежесткими моделями?

Прежде всего это создание реальной лингвистической среды для проведения экспериментов: и по отработке самой модели неполного понимания, и для пополнения, а также коррекции имеющихся Словарных сведений, и для накопления гипотез о разных «логических» хо-

дах авторов текстов и общих свойствах всех текстов... Все даже очевидные возможности такой среды не перечислить.

В дополнение к модным сейчас направлениям собирать корпуса однородных текстовых материалов я считаю необходимым создавать вычисляющие лингвистические комплексы. Любые текстовые корпуса нуждаются не столько в ручной обработке, сколько в стабильно работающих процессорах. Это должны быть **полные** системы, включающие, во-первых, совместимые средства автоматического анализа текстовых массивов, а во-вторых, аппарат оценки результатов анализа в каждом комплексе. Такая диагностика каждым уровнем анализа свойств создаваемого объекта очень ускорила бы работу по созданию динамичных экспериментальных моделей, подобных ИЛМ. Она может вестись параллельно с той серьезной и скрупулезной работой по теории семантики, которая ведется в нашей стране уже более полувека, снабжая нас ценнейшими лингвистическими (прежде всего лексикографическими) знаниями. Но одновременно возникает вопрос «Как можно использовать эти знания в Системе, если они даны тоже в виде ЕТ, и в лучшем случае как подробные и единообразные статьи энциклопедического словаря?»

Задача поиска содержательной информации в массивах ЕТ, а также другие, более серьезные задачи, как машинный перевод или автоматизация сбора знаний из текстов, требуют основательного лингвистического анализа, проходящего по всем значимым уровням и требующего современных методов программирования в реализации. Это достаточно трудозатратная деятельность. Полный семантический анализ произвольных текстов еще никем не реализован — это действительно задача века, как и создание полного Семантического словаря.

Поэтому экспериментальная модель ИЛМ имеет целью сделать процесс понимания текста динамичным и в то же время **мягким**, имеющим много «сменных» компонентов. У нее нет задачи строить один единственный (всегда один и тот же в данной системе) результат в виде жесткого семантического представления текста. В каждом сеансе работы системы ВУ может создавать свое информационное представление (или индивидуальное ИнфП), постепенно адаптируя инструменты Системы к своим потребностям и «обкатывая» сам процесс анализа.

Итак, СемП — это внутреннее представление содержания текста, ИнфП — его внешнее представление, отражающее то, как текст воспринимается внешней средой. Одному тексту можно сопоставить бесконечно много разных ИнфП — столько, сколько возможно разных его «прочтений» или отображений в разных базах данных, и множество СемП, соответствующих разным возможным степеням сжатия. Это естественно, поскольку результаты процесса понимания (как и описания значений лексем в Семантическом словаре), как правило, субъективны, неполны, они могут быть ошибочными, могут противоречить общественным знаниям, и т.д.

Обсуждаемая нами структура семантического пространства текста не так требовательна к точности и тонкости описания значений, как того требует формальная семантика. Структура СемПрост неполна, неоднозначна, рыхла, но имеет простейший и единообразный синтаксис того метаязыка, на котором можно представить любые источники знаний, как «правильные», так и имеющие разные естественные недостатки.

Изучение закономерностей этого слабоструктурированного объекта может дать неожиданные эффекты не только в изучении процесса понимания текстов и в обкатке лингвистического обеспечения. Может быть, оно позволит проследить некоторые проблемы социального свойства: способы формирования слоев общества и его расслоение, образование кланов, моды, логики и шаблонов поведения и др. Ведь жизнь слов в тексте подобна поведению людей в обществе. Но ограниченная только вербальным материалом структура гораздо доступней для изучения упомянутых и других явлений, чем реальная общественная жизнь со всеми ее противоречиями и субъективными суждениями о ней.

О формализации понятий в анализе неструктурированной информации

Майсурадзе Арчил Ивериевич,
Московский государственный университет
им. М.В. Ломоносова

Введение

В последнее время потребности в анализе неструктурированной информации постоянно растут. Запрос к специалистам в области информационных технологий на автоматизированную поддержку такого анализа идет непосредственно от аналитических центров, а также транслируется через социологические научно-образовательные центры.

На российском рынке программного обеспечения уже присутствует ряд решений. Однако практически все эти решения, особенно с конкурентоспособной поддержкой русского языка, можно отнести к классу платформ для развертывания промышленных систем. Наблюдается серьезная нехватка решений для поддержки индивидуальной работы аналитиков. Обзор 38 программных систем (практически все зарубежные) для анализа неструктурированной информации и визуализации результатов такого анализа, проведенный летом 2010 года в АЦ при Правительстве РФ, показал, что все они существенно не соответствуют пожеланиям индивидуальных пользователей. Представляется, что основное препятствие для развития рассматриваемого класса ПО методологическое.

Группа специалистов в области социологии и информационных технологий из МГУ и ВШЭ разрабатывает персональное автоматизированное рабочее место для анализа неструктурированной информации, основными преимуществами которого являются: развитие системы от содержательных задач предметной области (а не от формальных методов), формализация и тиражирование экспертного знания методами машинного обучения.

Предлагаемая вниманию читателей работа демонстрирует ряд особенностей и сложностей, возникших в ходе совместной работы специалистов разных дисциплин над развитием общей методологии. В частности, будут рассмотрены разные подходы к пониманию, формализации и решению задачи «выделения концептов».

Трудности междисциплинарного общения

При решении задач, определенных данными, изложенными на естественном языке, возникает целый комплекс проблем. Такие проблемы уже с середины 60-х годов фиксируются и обсуждаются в публикациях. Но только сравнительно недавно стало понятно, что одна из основных проблем состоит в том, что «естественнонаучные» специалисты, создающие программные решения, не знают, какие задачи на самом деле решают их «гуманитарные» потребители – эксперты в предметных областях. Конечно, поиск взаимопонимания является неотъемлемой частью любого междисциплинарного исследования. К сожалению, в области анализа неструктурированной информации границы между научными дисциплинами совпали с границами между производителями и потребителями ПО. В итоге, развитие индустрии проходит гораздо менее сбалансировано между предложениями и потребностями, чем в области анализа структурированной информации. Например, в области интеллектуального анализа (структурированных) данных лидеры рынка ПО уже предлагают системы для решения содержательных задач, четко декларируют, что их ПО поддерживает соответствующие содержательным задачам методологии и индустриальные стандарты; объекты анализа точно формализованы, алгоритмы, реализующие методы ИАД, четко описаны.

Программные системы для анализа неструктурированной информации до сих пор предоставляют, по сути, лишь некоторые библиотеки функций, причем список этих функций сформирован самим производителем. Обычно для сторонних специалистов (даже в области информационных технологий) остается непонятным, почему пользователю был предложен именно такой список функций. Обычно неясно, как именно формализованы понятия, используемые в определении таких функций. А когда речь заходит о методах, лежащих в основе функциональности, а тем более о точных реализациях алгоритмов, тут уж сами производители ПО стараются сказать как можно меньше. Например, в ходе круглого стола по вопросам

анализа неструктурированной информации в социологических исследованиях, прошедшего в ВШЭ зимой 2011 года, неразрешимым вопросом для поставщиков ПО оказался следующий: «Укажите публикации по реализованным в Вашем ПО методам, на которые можно сослаться в научных работах».

Указанное положение дел неудобно для потребителей ПО, а для научной деятельности оно практически неприемлемо, так как методология научных исследований не поощряет рассуждения, в которых «неясно что» сделано «не пойми как». Даже если авторы методов, реализованных в коммерческом ПО, где-то публикуются (например, можно упомянуть труды конференции по компьютерной лингвистике «Диалог»), отсутствует связь между четким описанием метода и конкретным программным продуктом.

Представляется, что описанная ситуация сложилась именно из-за недостаточного развития методологии использования автоматизированных программных средств для решения содержательных задач. Думается, что сложности развития такой методологии обусловлены тем, что вообще методологии естественнонаучного и гуманитарного исследования несколько отличаются. Если в естественных науках принято практически всегда проводить логические рассуждения с использованием точно формализованных «объективных» понятий, то в гуманитарных дисциплинах важную роль играет сравнение «субъективных» точек зрения.

Как показало личное наблюдение автора за практикой взаимодействия социологов и специалистов в области информационных технологий, последние не сразу понимают, что первые хотят сделать. И тогда вместо обмена идеями происходит лишь обмен словами; обе стороны пытаются ухватиться за некоторый набор терминов, которые, как им кажется, они совместно используют. Трудность не только в том, что в этот набор попадают термины, имеющие различное определение в разных дисциплинах. Серьезная проблема в том, что алгоритмисты вынуждены дополнять этот набор словами, которые социологи упомянули в формулировке задач. Специалисты в области естественных наук вынуждены строить свои определения таких понятий, пришедших из предметной области.

В настоящее время можно говорить о следующих основных подходах к формализации понятий предметной области.

- Специалисты в области естественных наук сами предлагают какую-то формализацию, практически без участия потребителей. Сегодня это наиболее распространенный подход в области анализа неструктурированной информации. В настоящее время основная проблема этого подхода состоит в том, что формализация понятий почти всегда идет от некоторого метода. Методологически должно быть наоборот: сначала ставится задача, а уж потом рассматриваются разные методы её решения. Нередки жалобы на то, что как алгоритмист не понял, чего хотел потребитель, так и потребитель не понял, что предложил алгоритмист.
- Формализация фиксируется как набор готовых ответов для заданного набора конкретных ситуаций. Профессиональная подготовка набора готовых ответов занимает годы. При появлении новых ситуаций потребителю придется довольно долго ждать поддержки со стороны производителя. Потребитель обязан согласиться с предоставленными ему готовыми ответами. Типичный пример – словари, тезаурусы.
- Просим у пользователей обучающие примеры. Много разных несогласованных мнений. Не всегда понятны причины мнения. Мнение нестабильно (меняется во времени). Мнение зависит от текущей деятельности. Эксперт не хотел бы тратить время на подготовку примеров. Способ узнать мнение у эксперта опять-таки требует формализации. Можно индивидуально поддерживать одновременно разные виды деятельности или разных экспертов. Оперативная актуализация системы.
- Извлекаем обучающие примеры из уже накопленной информации. В первую очередь речь идет о том, что в Интернет накоплено колоссальное количество примеров, которые можно использовать для обучения. Пока больше надежд, чем результатов. Подход дает возможность следить как за поставщиками информации, так и за её потребителями.

В настоящее время на рынке программных средств для анализа неструктурированной ин-

формации первые два подхода получили подавляющее распространение. В данной работе приводятся примеры работы в рамках подходов 3 и 4.

Задачи анализа текстов

Разумеется, когда речь идет об анализе информации, выраженной текстом на естественном языке, в самой общей постановке хотелось бы говорить о понимании машиной текста, автоматическом выделении смысла. В формулировке, приближенной к реальным возможностям современных систем, стоит задача восстановления отдельных объектов и их взаимосвязей, которые были описаны, либо упомянуты, либо подразумевались автором текста неявно. Это заставляет обращаться к работе с некоторыми объектами, отсутствующими в тексте в явном формализованном виде, но описанными автором и несущими реальное значение для решения содержательной задачи.

Рассмотрим некоторые примеры, в которых пытаются формализовать и решить задачу поиска смысловых понятий текста и взаимосвязей между ними.

В ходе построения информационно-поисковой Интернет-системы задача в первую очередь сводится к следующему. Машине необходимо составлять базу данных, в которой хранится информация о некоторых объектах, процессах и явлениях, описанных в текстах, и эти записи сопровождаются информацией о свойствах, качествах и взаимосвязях описанных в текстах объектов. При этом один и тот же объект может описываться с использованием различных слов (проблема использования синонимов). Объект может даже не описываться, а упоминаться косвенно, например, надо найти тексты о бейсболе, в которых слова «бейсбол» нет. Кроме этого, в различных текстах одинаковыми словами могут описываться различные объекты (проблема использования омонимов). В информационно-поисковой системе, с одной стороны, стоит задача сужения области поиска, исключения из нее документов, упоминающих ненужные пользователю объекты/события, а с другой стороны, возникает необходимость застраховаться от излишнего сужения, традиционно возникающего за счет того, что пользователь может спрашивать об объекте совсем не теми словами, которыми пользовался автор при описании.

Вслед за проблемой восстановления объектов, описанных в тексте, возникают проблемы восстановления взаимосвязей, отношений, характеристик и так далее. Если задача по выделению объектов в упрощенном виде решается известными средствами — составление словарей объектов, словарей синонимов и т.д., — то восстановление связей, отношений и характеристик, особенно описываемых неявно и/или разрозненно, является весьма сложной задачей, не решаемой без применения интеллектуальных технологий, базируемых на проработанных моделях естественного языка, моделях построения текстов, моделях мышления.

Заслуживает упоминания задача дешифровки исторических систем письма. Имеется осмысленный текст, записанный на неизвестном языке. Нужно, исходя в первую очередь из самого текста, выяснить свойства неизвестного языка и уже затем, путем сопоставления с известными языками и привлечения с большой осторожностью внетекстовой информации (археологических, исторических, филологических и иных сведений), передать смысл неизвестного текста. Возможность исследовать текст формальными методами перерастает в необходимость, если требуется максимальным образом исключить субъективный анализ текста.

Схожие задачи стоят и перед разработчиками программного обеспечения машинного перевода текстов. Похожей проблемой занимаются разработчики систем создания и определения неестественного происхождения документа. Упомянем задачи распознавания спама по контексту, задачу построения семантических сетей, используемых в словарях и тезаурусах, задачи рубрикации текста, выделения тематики, автоматического аннотирования и реферирования. Существуют также более локальные задачи, например, задача автоматического поиска тавтологий в тексте или задача автоматического переноса слов (например, в английском языке омонимы переносятся по-разному).

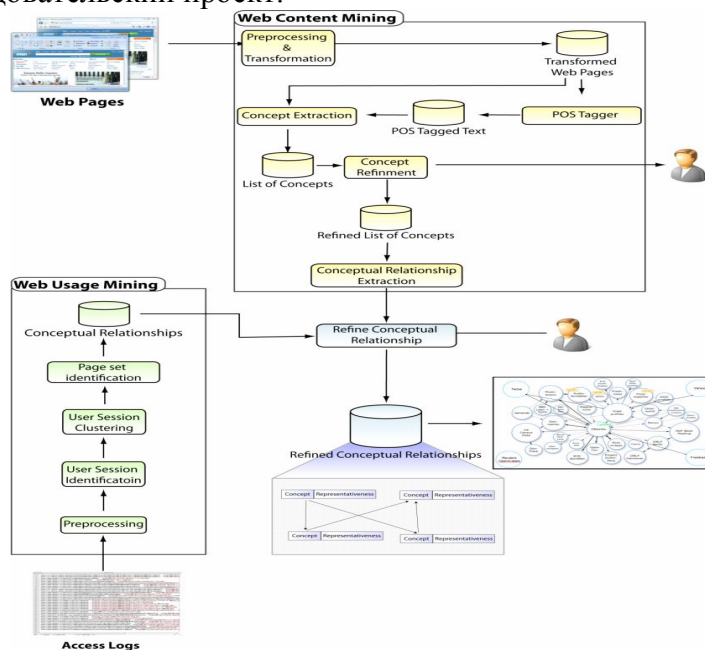
В последнее время стали появляться работы по автоматизации накопления и использования общих знаний об окружающем мире, необходимых для понимания текста. Кроме того, наметились подходы для анализа одного и того же текста, как с точки зрения автора, так и с

точки зрения читателя. В данной работе в качестве основного иллюстративного примера использована задача выделения концептов.

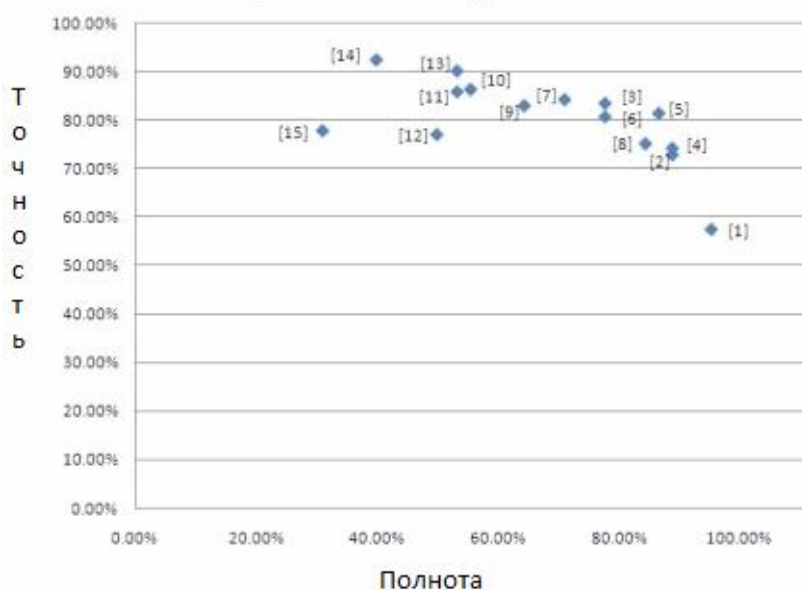
Использование накопленной информации

В настоящее время формализацию знаний о предметной области чаще всего реализуют в виде онтологий. Онтологии используются в процессе программирования как форма представления знаний о реальном мире или его части. Большинство используемых сегодня онтологий созданы вручную, их разработка и наполнение — это долгий и трудоёмкий процесс. Они оцениваются чаще с точки зрения применимости, чем полноты. Пользователь практически лишен возможности актуализировать накопленные в системе знания самостоятельно.

Ниже приводится архитектура системы, предназначенной для полуавтоматического наполнения онтологии информацией. Новизна системы определяется тем, что собирается как информация о текстах определенных авторов, так и информация, отражающая поведение читателей. Пример приводится именно для того, чтобы продемонстрировать сочетание двух потоков информации, а не для рассказа о методах. Указанная система разрабатывается как студенческий исследовательский проект.



Оценка качества выделения объектов



В данной работе основное внимание уделяется задаче выделения концептов, что в теории

онтологий соответствует экземплярам и понятиям. Экземпляры — это отдельные объекты, понятия (классы) — абстрактные группы объектов. Онтология может обойтись без конкретных объектов или не различать экземпляры и понятия. Поэтому позволим себе использовать только термин концепт.

Система сочетает в себе два подхода: Web Content Mining и Web Usage Mining. Первый подход анализирует текст и структуру страниц, именно эта информация характеризуют автора. Второй подход анализирует информацию о поведении читателей. Интерес представляет именно совместное применение подходов.

При анализе текста концепты выделяются на основе взвешенных частот. Использовать непосредственно частоты не удастся, так как размер каждого отдельного текста мал. Веса назначаются на основе HTML-тэгов, в которые концепт заключен. Кроме того, привлекается морфологическая информация, в особенности о частях речи. Это полезно для правильного отбора цепочек слов, обозначающих концепт.

Информация об использовании страниц позволяет выявить шаблоны поведения читателей. Разумно предположить, что пользователи перемещаются по сайту в соответствии со своим восприятием предлагаемого материала. Такой анализ группирует страницы, даже если между ними нет прямых связей.

Постараемся оценить качество результатов, даваемых системой. Трудность в том, что для онтологий не выработано стандартных показателей качества. Ближе всего в данной ситуации представляются понятия точности и полноты, используемые при оценке качества информационного поиска. Однако при этом приходится проводить сравнение результатов работы системы с некоторой эталонной онтологией. Результаты нескольких запусков системы приведены на диаграмме.

Использование экспертных примеров

В данном разделе рассматривается задача выделения «концептов» и «релевантных понятий» в одном тексте. Проведем частичную формализацию терминов «концепт» и «релевантное понятие». Концепт в тексте представлен одним или несколькими вхождениями слов или словосочетаний. Концепт — это некоторый набор вхождений; релевантные понятия — такие концепты, которые «соответствуют ключу информационной сферы». Такая частичная формализация была положена в основу опроса экспертов, чтобы собрать примеры «концептов» и «релевантных понятий». Было принято решение отобрать ряд русских текстов и предложить группе экспертов выделить в них группы вхождений, соответствующие концептам и релевантным понятиям.

Отбор текстов осуществлялся на основе следующих посылок. Во-первых, эксперты должны с текстами справиться, а значит, объем каждого текста и их количество в опросе должны быть ограничены. Несколько попыток показали, что для человеческого восприятия удобнее всего маленькие, порядка 1000 символов, тексты — это два-три небольших абзаца, или один большой. Во-вторых, каждый текст должен быть связным и завершенным, в нем должна четко прослеживаться основная идея. Это не должен быть «выданный из контекста кусок». Лучше всего поставленным требованиям отвечают тексты в жанре журналистских заметок. Это своего рода резюме к большой статье, где акцент сделан уже на основных ключевых моментах текста, а потому экспертам довольно легко выделить искомые понятия.

Тематика для текстов, входящих в набор, выбиралась одна: «Последствия аномальной жары и лесных пожаров в России летом 2010 года». Это связано с тем, что наличие одной проблемы в корпусе практически наверняка гарантирует составление одного терминологического словаря. Такой стиль исследования называется работой в одном дискурсе. К тому же, выбор темы был обусловлен сравнительной простотой и равной актуальностью проблемы для всей экспертной группы.

Некоторые характеристики набора текстов. Максимальное число символов в одном тексте 1787, среднее число символов 1060, минимальное число символов 646. Максимальное число слов в тексте 231, среднее число слов 137, минимальное число слов 85. Максимальное число предложений в тексте 20, среднее число предложений 8, минимальное число предложений 3.

Максимальное число абзацев 4, среднее число абзацев 3, минимальное число абзацев 2.

Выбор экспертной группы основывался на следующих факторах. Каждый эксперт должен воспринимать тематику текстов, понимать, что требуется найти в предложенном тексте, иметь начальные профессиональные навыки обработки текстов с лингвистической точки зрения. В частности, перечисленным требованиям удовлетворяют люди с филологическим или лингвистическим образованием. В качестве экспертов были привлечены студенты факультета иностранных языков и регионоведения МГУ им. М.В. Ломоносова.

В данной работе после опроса экспертов мы хотим получить концепты как группы, состоящие из одного или нескольких вхождений, и помеченные среди них особые группы, которые мы называем релевантными. При обдумывании вопроса, в каком виде мы бы хотели получить информацию от экспертов, возник другой важный вопрос: а как экспертам было бы удобнее (интереснее) решать поставленные перед ними задачи. Оказалось, что заполнять какую-либо анкету по графам эксперты не хотят.

В результате нескольких проб сформировалась следующая процедура опроса. Экспертам предлагалось помечать вхождения, относящиеся к разным концептуальным группам, фломастерами разных цветов, а рядом с группой, которую они считают релевантной, ставить какой-нибудь отличительный символ, например, звездочку. На практике эксперт крайне редко выделял в одном тексте более 4 групп, так что количество доступных фломастеров (не менее 8) не было ограничением.

Дополнительно в анкете, предложенной экспертам, была отдельная графа под комментариями. В этой графе каждый эксперт мог в свободной форме указать, по какой причине он сгруппировал те или словосочетания, почему считает выделенные им группы релевантными, попытаться подумать о тех закономерностях, на которые можно было бы впоследствии ориентироваться при составлении алгоритмов.

Как показывает практика, даже с маленькими текстами экспертам утомительно работать, если их много. Поэтому было решено каждому эксперту предъявлять по 8 текстов. Каждый текст был предъявлен хотя бы 4 экспертам. Эксперты работали над текстами от 40 минут до полутора часов. Приведем результаты опроса на примере одного из текстов.

«Реальная площадь лесных пожаров с начала пожароопасного периода 2010 года в России примерно вдвое превышает данные субъектов РФ и составляет около трех миллионов гектаров, сообщил в понедельник журналистам начальник Авиалесоохраны Федерального агентства лесного хозяйства России Николай Ковалев.

По официальным данным Рослесхоза, на начало сентября текущего года во время пожароопасного периода на территории страны возникло более 30 тысяч природных пожаров общей площадью более чем 1,246 миллиона гектаров. Однако, эти данные по площади лесных пожаров, предоставленные регионами, не совпадают с данными аэрокосмического мониторинга, проведенного в то же время, и сильно занижены. Это связано с тем, что главы ряда субъектов не хотят показывать истинное положение дел».

№ эксперта	Группы концептуальных понятий	Релевантные понятия	Комментарии
1	<u>Группа 1:</u> Реальная площадь лесных пожаров (1, 1–4), истинное положение дел (2, 98–100); <u>Группа 2:</u> По официальным данным Рослесхоза (2, 38–41), данные по площади лесных пожаров, предоставленные регионами (2, 69–75).	Группа 1	Концептуальными группами являются те, ядром которых служит существительное.
2	<u>Группа 1:</u> Реальная площадь (1, 1–2), на территории страны (2, 51–53), общей площадью (2, 60–61), по площади (2, 70–71); <u>Группа 2:</u> лесных пожаров (1, 3–4), пожароопасного	Группа 1 Группа 4	Группа 1 — мера площади и пространства Группа 2 — пожары и

	периода (1,7-8), пожароопасного периода (2, 49-50), природных пожаров (2, 58-59), лесных пожаров (2, 72-73); <u>Группа 3:</u> 2010 года (1, 9-10), трех миллионов (1, 22-23), 30 тысяч (2, 56-57), 1,246 миллиона (2, 64-65); <u>Группа 4:</u> вдвое превышает (1, 14-15), данные (1, 16), не совпадают (2, 76-77), сильно занижены (2, 88-89), не хотят показывать истинное положение дел (2, 98-103).		все что их касается Группа 3 — по наличию цифр Группа 4 — несовпадение предоставляемой информации с реальностью
3	<u>Группа 1:</u> Реальная площадь (1,1-2), около трех миллионов гектаров (1,21-24), площадью более чем 1, 246 миллиона гектаров (2,61-66); <u>Группа 2:</u> лесных пожаров (1,3-4), пожароопасного периода (1,7-8), пожароопасного периода (2,49-50), природных пожаров (2,58-59), лесных пожаров (2,72-73).	Группа 1	
4	<u>Группа 1:</u> Реальная площадь лесных пожаров (1, 1-4), составляет около трех миллионов гектаров (1, 20-24), на территории страны (2, 51-53), 30 тысяч природных пожаров общей площадью более чем 1,246 миллиона гектаров (2, 56-66), данные по площади лесных пожаров (2, 69-73); <u>Группа 2:</u> с начала пожароопасного периода 2010 года (1, 5-10), на начало сентября текущего года (2, 42-46).	Группа 1	Группа 1 — площадь Группа 2 — время

Как и ожидалось, ответы экспертов частично разошлись. Интересно отметить, что некоторые группы сформированы по составу, например, по наличию цифр. Важно подчеркнуть, что кроме общей идеи выделения групп, сформулированной в начале раздела, никакое мнение экспертам не предлагалось. В некоторых случаях эксперты составили такие группы, о которых авторы исследования не предполагали. Следовательно, если бы алгоритмисты попытались сами формализовать задачу или метод решения, то неизбежно работали бы с неверным представлением о потребностях потребителей.

Реальная площадь лесных пожаров с начала пожароопасного периода 2010 года в России примерно вдвое превышает данные субъектов РФ и составляет около трех миллионов гектаров, сообщил в понедельник журналистам начальник Авиалесоохраны Федерального агентства лесного хозяйства России Николай Ковалев.

По официальным данным Рослесхоза, на начало сентября текущего года во время пожароопасного периода на территории страны возникло более 30 тысяч природных пожаров общей площадью более чем 1,246 миллиона гектаров. Однако, эти данные по площади лесных пожаров, предоставленные регионами, не совпадают с данными аэрокосмического мониторинга, проведенного в то же время, и сильно занижены. Это связано с тем, что главы ряда субъектов не хотят показывать истинное положение дел.

Был предложен способ визуализации результатов опроса, названный картой концептов. Работе 4 экспертов соответствуют цвет, фон, размер и эффект шрифта.

Для применения методов ИАД в дальнейшем была сформирована обучающая выборка, для чего консолидация мнения экспертов проводилась двумя способами: наложение групп и их разбиение. В итоге каждому тексту соответствовало несколько групп с разными весами.

Дискурс-анализ электронных СМИ с применением сетевого подхода (пример обсуждения вступления РФ во Всемирную торговую организацию)

Просьянюк Дарья Вячеславовна, *НИУ ВШЭ*

Сегодня каждое принятое или готовящееся к принятию экономико-политическое решение находит отражение в информационном Интернет-пространстве. Здесь каждая страна имеет возможность собственного позиционирования и коммуникации с другими странами. Поэтому планирование долгосрочной идеологической перспективы действий в этом полифакторном пространстве представляется вопросом исключительной важности. Примером такого экономико-политического решения может служить вступление Российской Федерации во Всемирную Торговую Организацию. Данная проблема уже не первый год является крайне актуальной для российской внутренней и внешней политики и экономики.

Любое экономико-политическое решение (особенно такого масштаба как вступление в ВТО) должно получить легитимацию в различных слоях населения. Соответственно, информационное пространство — это визуализатор экономико-политических решений, которые принимаются и готовятся к принятию, это своего рода «поле», где взаимоотношения между странами становятся очевидными. Таким образом, информационное пространство является уникальным репрезентатором внутри- и межстрановых отношений: в нем апробируются способы аргументации и легитимации принимаемых «наверху» решений, а также с относительно большей полнотой, чем в других открытых источниках, эксплицируются неявные взаимоотношения и взаимосвязи между странами.

1. Методика

Шаг 1: выбор объекта. Всякий социально-значимый объект описывается в информационном пространстве в различных аспектах. На первом этапе построения эгоцентричной сетевой карты выбирается аспект, который за счет представленности в различных источниках информации обеспечит достаточный размер выборки и вариативность типов информационных сообщений. В нашем исследовании для построения эгоцентричной сетевой карты была выбрана тема «Вступление Российской Федерации во Всемирную торговую организацию».

Шаг 2: кодирование сообщений. Под сообщением мы понимаем всякий текст, имеющий отчетливо обозначенные границы. Традиционно в журналистике принято выделять три жанра написания текстов: информационный, аналитический и художественно-публицистический¹. Для построения эгоцентричной сетевой карты используются тексты информационного и аналитического жанров. Характер содержания в текстах информационной группы эмпирический: они в первую очередь передают аудитории фактические сведения о действительности. Такие тексты позволяют выявлять и классифицировать информационные поводы и информационные события. В аналитических текстах содержится информация эмпирико-теоретического характера: здесь главным образом выполняется задача систематизации фактов, их объяснения и обобщения, в них содержатся реконструкции позиций и типов аргументации, которых могут придерживаться различные представители разных стран. Описанные группы относятся к профессиональным текстам.

При построении эгоцентричной сетевой карты могут представлять интерес и непрофессиональные тексты — блоги. Поскольку они репрезентируют неофициальный взгляд на проблему, интерпретация информационных поводов в них может значительно отличаться от официальных источников.

¹ Ким М.Н. Жанры современной журналистики. М.: Изд-во Михайлова В.А., 2004.

Для кодирования информации разрабатываются специальные кодировочные макеты. Обычно макеты содержат формальное (например, дата публикации, название, страна, название издания, количество символов и пр.) и содержательное (это может быть, например, частота упоминания персон или организаций, наличие речевых маркеров, аргументации и пр.) описание статьи

Шаг 3: расчет весов аргументов и оценок. Заполненные базы данных содержат определенное количество аргументов, используемых авторами для обоснования своей точки зрения на обсуждаемую проблему. Предположим, во всем массиве статей трех стран А, Б и В было высказано 3 аргумента *a*, *b* и *в*. Сначала подсчитывается количество употреблений каждой страной каждого аргумента: допустим, страна А употребила аргумент *a* 1 раз, аргумент *b* — 2 раза, аргумент *в* — 3 раза, страна Б употребила аргумент *a* 0 раз, аргумент *b* — 1 раз, аргумент *в* — 1 раз; страна В употребила аргумент *a* 3 раза, аргумент *b* — 2 раза, аргумент *в* — 1 раз.

Затем высчитывается **доля каждого аргумента во всем дискурсе**: аргумент *a* был употреблен 4 раза (1 раз страной А, ни разу страной Б и 3 раза страной В), аргументы *b* и *в* по 5 раз, то есть доля аргумента *a* равна 0,29 (4/14), аргументов *b* и *в* — 0,36.

При умножении **доли каждого аргумента во всем дискурсе** на **частоту употребления аргумента каждой страной** получаем **взвешенную частоту использования страной каждого аргумента**. Данный показатель является вспомогательным для расчета оценки страны на основании аргументов.

В данном случае для страны А аргумент *a* = $1 \cdot 0,29 = 0,29$; аргумент «2» = $2 \cdot 0,36 = 0,72$; аргумент «3» = $3 \cdot 0,36 = 1,08$.

Затем высчитывается среднее арифметическое по всем аргументам для каждой страны. Для страны А оно составляет $(0,29 + 0,72 + 1,08) / 3 = 0,7$ — это и есть **оценка страны по данному вопросу на основе аргументов**. Данный показатель позволяет рассчитывать значение оценки страны, учитывая взвешенную частоту использования аргументов.

Отдельно учитывается оценка мнения страны по обсуждаемому вопросу: мнение может быть высказано (предположим, «России выгодно/невыгодно членство в ВТО»), тогда страна получает +/- 1 балл, или не высказано (то есть присутствует только аргументация без сформулированного мнения), тогда страна получает 0 баллов. Для подсчета интегрированной оценки страной обсуждаемого вопроса эти две величины складываются:

Вес аргумента по стране = частота его употребления этой страной * долю этого аргумента от всех аргументов.

Оценка страны обсуждаемого вопроса = среднее арифметическое весов аргументов + высказанная оценка.

Рассчитанная таким способом интегрированная оценка позволяет учитывать как прямую высказанное мнение, так и мнение, выявленное на основе приведенной аргументации. Необходимость такого подхода объясняется тем, что анализируемый текст может не содержать прямо вербализированной оценки/мнения/отношения автора к описываемому объекту/факту/событию, а включать ряд аргументов за/против. Предложенный способ расчета позволяет выявить и измерить интегрированную оценку обсуждаемого вопроса, эксплицированную в использованной аргументации.

Шаг 4: построение и анализ эгоцентричной сетевой карты. Эгоцентричная сетевая карта представляет собой графическое изображение (граф) межстранового информационного пространства, отражающее мнение стран по исследуемым вопросам.

2. Пример построения эгоцентричной сетевой карты: вступление РФ в ВТО

В качестве примера приведем эгоцентричную сетевую карту дискуссий в электронных СМИ о вступлении Российской Федерации во Всемирную торговую организацию.

Первоначально поиск статей проводился на следующих ресурсах:

— www.inosmi.ru, www.inopressa.ru — ресурсы специализируются на переводе на русский язык материалов зарубежных СМИ о России;

— www.wto.ru — сайт публикует информационные и аналитические статьи о ходе продвижения переговоров о вступлении РФ в ВТО, официальные документы ВТО;

— поисковые системы «Yandex» и «Google» (поиск по блогам) — ведущие поисковые системы, обрабатывающие 46 и 34 % поисковых запросов Рунета соответственно¹; поиск производился по ключевым словам «Российская Федерация и Всемирная торговая организация» («Russian Federation and World Trade Organization»).

Таким методом был отобран корпус статей и составлен список изданий, опубликовавших хотя бы одну статью по теме исследования за указанный период. Затем поиск проводился на сайтах изданий из сформированного списка, а также правительств соответствующих стран. Были отобраны все статьи, обсуждавшие вступление России в ВТО за период с января 2008 по январь 2010 г.: из 50 источников 15 зарубежных стран была отобрана 61 статья, из 85 российских источников — 163 статьи.

Эгоцентричная сетевая карта

Эгоцентричная сетевая карта — графическое изображение (граф) межстранового информационного пространства, репрезентирующее средневзвешенную оценку странами трех основных вопросов: оценку шансов вступления РФ в ВТО, оценку перспектив сотрудничества с Казахстаном и Белоруссией (таможенное соглашение) и оценку мнения РФ по поводу целесообразности вступления в ВТО (см. рис.).

Как видно, большинство рассматриваемых стран негативно оценивают шансы вступления РФ в ВТО. Особым скептицизмом отличается США, Литва, Грузия, Франция, Пакистан и Азербайджан, далее следуют Финляндия, Бельгия, Армения и Польша, чуть менее пессимистичны Великобритания и Латвия. Для Германии и Украины характерна нейтральная позиция: примерно равное количество аргументов «за» и «против» и нет однозначной оценки. Оптимистами являются лишь эстонцы (которые видят выгоду для себя от нашего вступления в ВТО и, возможно, выдают желаемое за действительное) и наши партнеры по таможенному союзу белорусы.

Отношения стран к перспективам нашего сотрудничества в Белоруссии и Казахстане в рамках таможенного союза более спокойные: здесь больше нейтральных позиций (Германия, Украина, Армения, Беларусь, Грузия, Латвия, Пакистан, Польша, Эстония), а негатив (Великобритания, США, Бельгия, Литва, Финляндия) выражается более мягкими оценками. Возможно, это объясняется тем, что, несмотря на негативное влияние ТС на возможность вступления РФ в ВТО, он может принести пользу и укрепить экономики формирующих его стран (особенно в посткризисном мире).

Что касается оценки позиции РФ относительно целесообразности вступления в ВТО, то здесь лишь две страны верят в желание Кремля присоединиться к ВТО в скором времени: Украина (которая считает, что вступления в ВТО — средство политического соперничества с ней РФ) и Финляндия (думающая, что ВТО — это желаемый толчок для российской экономики).

Сама же Россия не надеется на скорое вступление в ВТО (общая оценка –0,2), не приветствует создание таможенного союза (общая оценка –0,2), а по поводу целесообразности вступления приводит множество аргументов «за» и «против», которые позволяют сделать вывод о том, внутри страны есть заинтересованные группы, мнения которых в некоторых вопросах прямо противоположны. Так что однозначного вывода о выигрыше или проигрыше РФ при вхождении ее в ВТО сделать нельзя.

Далее, на основе позиций стран по трем основным вопросам был проведен иерархический кластерный анализ методом межгрупповых расстояний. По его результатам было решено выделить три кластера. В первый кластер «настороженных оптимистов» вошли Украина и Финляндия, которые с пессимизмом оценивают шансы вступления РФ в ВТО, с опаской относятся к созданию ТС, но верят в желание России скорее примкнуть к рядам этой многосторонней организации. Самый многочисленный второй кластер пессимистов составили Великобритания, США, Бельгия, Франция, Азербайджан, Армения, Грузия, Литва, Латвия, Пакистан, Польша; они однозначно негативно оценивают все три возможности. Наконец, третий

¹ По данным доклада маркетинговой группы Newmann Bauer: Обзор и исследование поисковых систем. 24 июня 2010 г. <http://www.nbmarketing.ru/poisksystem/>

кластер оптимистов — Россия, Эстония, Беларусь, Германия — оптимистичны в оценках настроений РФ и ее шансов скорого вступления в ВТО, а также нейтрально оценивают перспективы создания ТС.

Анализ аргументации

Разработанные макеты для кодирования статей позволили выявить и проанализировать сопутствующую аргументацию позиций стран на три принципиальных вопроса: оценку шансов вступления РФ в ВТО, оценку перспектив сотрудничества РФ с Казахстаном и Беларусией в рамках таможенного союза и оценку мнения самой России по поводу целесообразности вступления в ВТО. Всего было выявлено около 50 аргументов по всем трем вопросам с российской стороны и примерно столько же с противоположной стороны. Из них совпали 14 % аргументов.

В аргументациях обеих сторон присутствуют как внутренние (нежелание самой России входить в ВТО), так и внешние причины негативной оценки шансов на вступление. При этом совпадающим является внешний, а, следовательно, объективный для России фактор (поправка Джексона-Вэника¹). Противоположными являются указание Россией на неприемлемые внешние требования, в то время как другие страны указывают на ее собственное нежелание вступать в ВТО. Также расходятся взгляды на желания западных стран: Россия отмечает их интерес к своему перспективному рынку, мир — его отсутствие.

Заявив в июне 2009 г. о своем намерении вступать в ВТО только в составе ТС, Россия, ожидая реакции, вынуждена была легитимировать данное решение. Так, были высказаны мнения о пользе такого объединения не только для экономик входящих стран, но и для перспектив их присоединения к ВТО. Запад также нашел в ТС некоторые плюсы, однако их список гораздо скромнее. В чем наши точки зрения совпадают, так это в том, что ТС — это хороший козырь в переговорах с западными партнерами. В оценке минусов мы более синхронны: здесь и формальная сложность образования ТС, и отсутствие прецедентов такого вступления. Кремль обосновывал перспективы такого образования модернизацией экономик, единым рынком от Польши до Китая, способом восстановления после кризиса. В зарубежных источниках никаких выгод выявлено не было. Аргументы о преимуществах объединения для локальных экономик скромны: Москва ожидает, что соседи будут делать так, как она хочет, Россия заботится о соседях.

Что касается оценки Россией собственного мнения о целесообразности вступления в ВТО, то преимущества и потери описываются с равной частотой. При этом преимущества носят обобщенный, теоретизированный характер: увеличение экспорта, модернизация экономики, объединение единомышленников (единственное исключение — невозможность экспортировать высокотехнологичную продукцию вне ВТО). Потери описаны конкретнее: ВТО не регулирует торговлю энергоносителями (составляющими две трети нашего экспорта), незаинтересованность государства и бизнеса, угроза многим отраслям (ВПК, топливно-энергетический комплекс, сельское хозяйство, автопроизводители, химическая промышленность, шины, чай, кофе, косметика).

Для западной точки зрения характерна констатация нежелания России вступать в ВТО (от недовольства бизнеса, протекционизма государства, формальной «неаккуратности» и недоверия таким организациям) и вера в благостное влияние ВТО на нашу экономику.

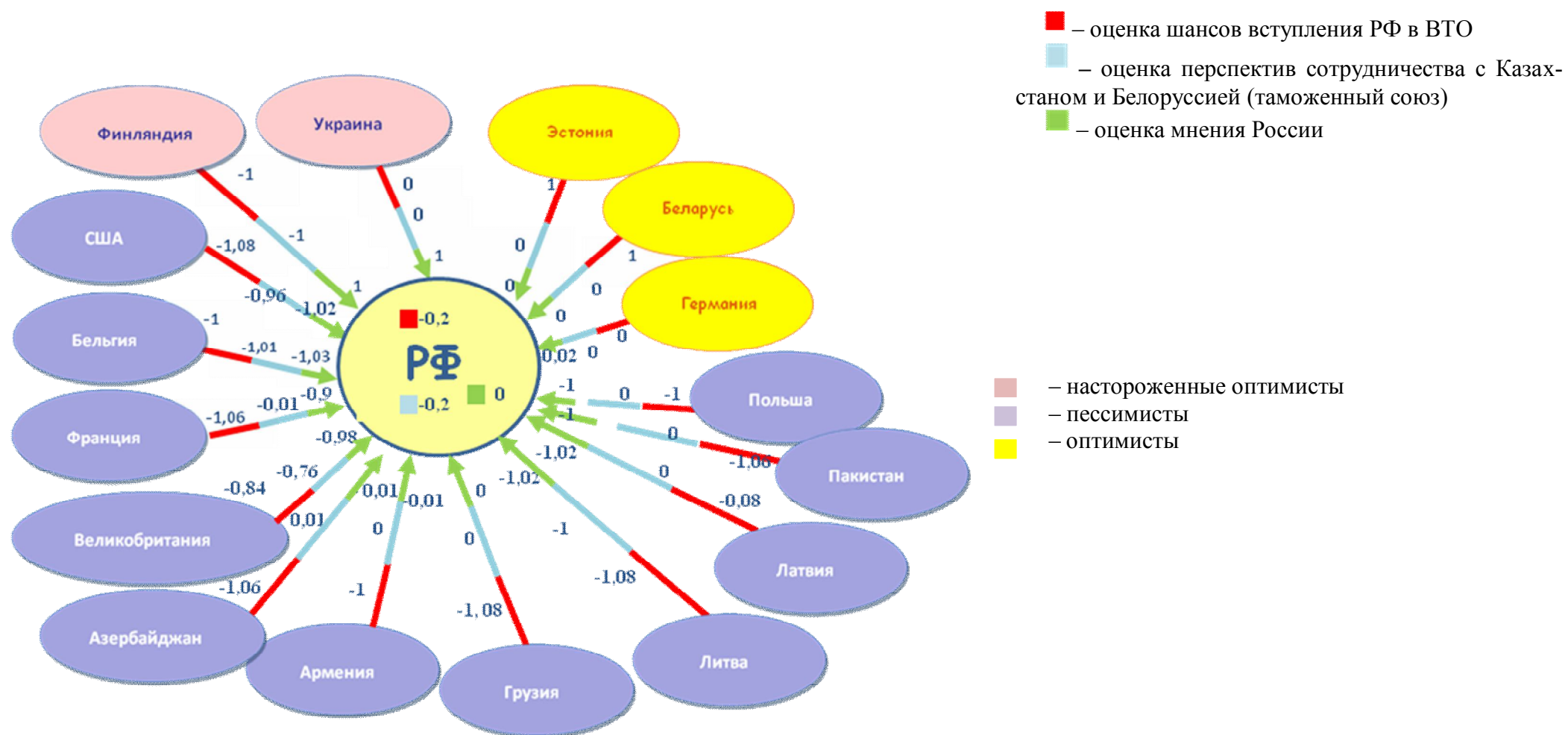
Прямо противоположных аргументов по данному вопросу выявлено не было, сходимся же мы в том, что членство в ВТО — это вопрос престижа, который чисто теоретически может помочь нашей экономике (в частности, прийти в себя после кризиса), но практически не отразится на основных статьях нашего дохода — нефти, уране, газе и оружии.

3. Заключение

В современном мире любое экономико-политическое решение нуждается в легитимации. Уникальный способ для этого — информационное Интернет-пространство, значение которо

¹ Поправка 1974 г. конгрессменов Г. Джексона и Ч. Вэника к Закону о торговле США, ограничивающая торговлю со странами социалистического блока, препятствующими эмиграции евреев и других своих граждан.

Рис. Эгоцентричная карта отношения стран мира к возможности вступления РФ во Всемирную торговую организацию



го в экономической и политической жизни любой страны трудно переоценить. Проблема вступления Российской Федерации во Всемирную торговую организацию — яркий пример тому. Являясь остро актуальной на протяжении уже многих лет, она находит свое отражение в этом пространстве.

3. Заключение

В современном мире любое экономико-политическое решение нуждается в легитимации. Уникальный способ для этого — информационное Интернет-пространство, значение которого в экономической и политической жизни любой страны трудно переоценить. Проблема вступления Российской Федерации во Всемирную торговую организацию — яркий пример тому. Являясь остро актуальной на протяжении уже многих лет, она находит свое отражение в этом пространстве.

В данной работе для анализа информационного Интернет-пространства и построения эгоцентричной сетевой карты использованы методы сетевого анализа, а также дискурс- и контент-анализ. Проведенная работа имеет большой потенциал сферы применения. Посредством подобной методики представляется возможным фиксировать взгляды ведущих стран и экспертов на ситуацию в мире, указать наличие основных интеллектуальных центров, в которых идет обсуждение мировых проблем. Посредством этих когнитивных карт, как нам представляется, можно будет отслеживать новые концептуальные модели развития политических и экономических событий, фиксировать наличие взаимосвязей между ведущими интеллектуальными центрами и центрами принятия важнейших экономико-политических решений.

Современные подходы к методологии контент-аналитических исследований

Семенова Анна Валерьевна,
Институт социологии РАН

Традиционный контент-анализ часто разделяется на явный и латентный контент-анализ. Явный контент-анализ направлен на наиболее очевидные и четко обозначенные значения текста, тогда как латентный контент-анализ распознаёт тонкие семантические нюансы сообщения, имеет дело со значениями более тонкими, неявными. Явный и латентный контент-анализ направлены на различные аспекты текста, но оба они используют ту же самую традиционную контент-аналитическую методологию для проведения исследования. Исследователями уже давно признавалось, что традиционный контент-анализ представляет особые трудности, когда применяется к «скрытому содержанию»¹. Эти трудности сосредоточены вокруг процедуры кодирования и обучения кодировщиков. Они проявляются в низком уровне совпадаемости результатов кодирования разными кодировщиками из-за их расхождений в понимании того, как именно текст должен быть закодирован. Исключительная трудность задачи соответствия требованиям обоснованности и надежности результатов (а именно, достижения приемлемого уровня согласия между кодировщиками) стала очевидна ещё в 1930–40-е годы, когда пионеры контент-анализа пытались продвинуть свою технику за пределы эксплицитного содержания сообщений. Эти проблемы латентного контент-анализа кроются в сложности его аналитических категорий. «Использование сложных по структуре и содержанию категорий создаёт серьёзные проблемы в отношении надёжности. Это проблема баланса между надёжностью процедур, с одной стороны, и богатством категорий, с другой. Какой смысл в том, что мы получаем надёжность, если в процессе мы утрачиваем важные оттенки смысла? Это критическая область для контент-анализа, и сочетание достоверности со сложными категориями все еще нуждается в адекватной проработке».²

Берельсон написал это более 50 лет назад, но проблема «все еще нуждается в адекватной проработке». Однако наука не стоит на месте и одна из важнейших проблем, занимающих аналитиков содержания, — модернизация методов. Одним из способов развития современной методологии является попытка ликвидации пропасти между количественными и качест-

¹ Laswell, H.D. The Technique of Symbol Analyses (Content-analysis). Experimental Division (1941)

² Berelson, Bernard. Content Analysis. In Communications Research, Glance (1952).

венными методами сбора и анализа данных. Речь идёт не о полном синтезе этих двух направлений, т.к. это пока не представляется возможным, но о частичном их совместном использовании на разных стадиях исследования.

Усложнение исследовательских задач в сфере коммуникационных исследований, явные выгоды комбинирования количественных и качественных методов логически привели исследователей к мысли о том, что включение в традиционно количественные методы (в частности, в контент-анализ) качественных компонентов и разработка на этой основе новых модификаций этого метода, широко используемого в коммуникационных исследованиях, даст исследователям возможность решать новые, более сложные задачи, расширить проблемное поле исследований, получать качественно новые, более глубокие и разнообразные выводы о природе изучаемых явлений. Очень интересный и перспективный вариант в этом плане был предложен на суд научной общественности Аароном Ахувиа в статье *Traditional, Interpretive, and Reception Based Content Analyses: Improving the Ability of Content Analyses to Address Issues of Pragmatic and Theoretical Concern*, Aaron Ahuvia/Social Indicators Research (May 2001), 54(2), pp.139-172.

Автор вышеназванной статьи предлагает изменить наш обычный взгляд на контент-анализ и приводит два варианта модернизации этой методологии. Он предлагает рассматривать термин контент-анализ не в традиционном смысле, как объективный, систематичный и количественный метод изучения содержания текстов, а как более широкое понятие, как методологию, с помощью которой текст кодируется по категориям, а затем подсчитывается частота появления каждой категории. Это более широкое определение позволяет предположить, что контент-анализ хотя и имеет количественный характер, но оставляет поле для дискуссии в отношении «объективности» (возможно более точный термин «повторяемость») и систематизации. Автор считает, что если раньше этот метод использовался как метод количественного определения содержания текста, то теперь мы можем пойти дальше и использовать его как метод подсчёта интерпретаций. На этой основе предлагаются две новых разновидности контент-анализа. *Reception based content analyses* (контент-анализ, основанный на ментальном восприятии или рецептивный) позволит исследователям просчитать то, как разная аудитория поймёт текст. *Interpretive content analyses* (интерпретативный контент-анализ) создан специально для латентного (скрытого) анализа содержания, в котором исследователи идут дальше определения количества прямых значащих элементов текста. Автор не призывает отказаться от тех основ, что стали стандартным контент-аналитическим подходом, но стремится сделать следующий шаг и выйти за рамки принципа «один размер соответствует всему». Предлагаемые им подходы весьма гибкие. Их реальное воплощение в каждой конкретной ситуации может иметь свои особенности и комбинировать (объединять) некоторые аспекты каждого из методов.

Интерпретативный контент-анализ

Итак, интерпретативный контент-анализ был предложен как метод для исследования латентного содержания и решения других сложных задач кодирования. Чтобы понять принципы, на которых основан интерпретативный контент-анализ, необходимо чётко определить различия между явным и латентным содержанием. В этом может быть полезна семиотическая теория денотативных и коннотативных значений (Berelson 1952). Обозначающие или денотативные (*denotative*) значения, относятся к явному содержанию и идентифицируют части текста. Они являются значениями первого порядка и отвечают здравому смыслу, имеют очевидное значение. Созначающие или коннотативные (*connotative*) значения (латентное содержание) реализуются через объединение (комбинацию) смыслов индивидуальных элементов текста с пониманием его смысла как целого.

Лассуэл проводит различие между латентным и явным содержанием, называя латентное содержание «интерпретацией», подразумевая, что явное содержание в любом случае не является интерпретацией, оно однозначно.¹ Автор же статьи считает, что и явное, и латентное со-

¹ Laswell H.D. The Technique of Symbol Analysis (Content Analysis), Experimental Division (1941).

держание сути интерпретации. В Вебстеровском словаре интерпретация определяется как «приписывание значения абстрактным символам». «Всякий раз, когда мы смотрим на черные значки на листе бумаги и распознаем их как значимое слово, мы осуществляем интерпретацию» - утверждает автор новых методик. Но денотативная интерпретация настолько для нас обычна и часто практикуема, что мы осуществляем ее, не осознавая, что это акт интерпретации. Это может создавать иллюзию того, что денотативные значения, как мы их воспринимаем, это физические части текста, а не интерпретации. Тем не менее, и явный, и латентный контент-анализ – это формы семантического анализа. А это говорит о том, что они оба осуществляют интерпретацию значений. Этот факт важен, т.к. даёт возможность усовершенствовать применение контент-анализа в отношении коннотативных значений.

Мы знаем, что традиционный контент-анализ может быть разделён на три стадии: выбор текстов, кодирование текстов и интерпретация результатов кодирования. Именно во время кодирования коннотативные категории создают проблемы. Эти проблемы возникают потому, что для коннотативных интерпретаций:

- 1) кодирование не может полностью замыкаться набором правил кодировки;
- 2) может значительно возрасти время обучения кодировщиков, из-за необходимости разбора всё новых возникающих вариантов интерпретаций при кодировании или даже необходимый уровень обучения может быть вообще не достижим;
- 3) правильное выполнение интерпретаций требует «теоретической чувствительности», которой не каждый обладает в равной мере, но которая может максимизироваться через сотрудничество в большей степени, чем через независимую работу кодировщиков.

С самого начала своего возникновения и распространения традиционный контент-анализ подразумевал разработку и использование формальных правил кодирования, увеличивающих совпадаемость результатов кодирования, полученных от разных кодировщиков. Однако использование правил кодирования было большим, нежели просто способом достижения согласия между кодировщиками. Это было процедурой, базирующейся на теории, описывающей процесс извлечения людьми значений из текста. На начальном этапе развития контент-анализа в социальных науках был популярен Парсоновский взгляд на проблему: люди следуют правилам интерпретации, которые позволяют им находить значения.¹ Если процесс нахождения значений в тексте по сути своей есть результат следования правилам, то кажется естественным и уместным сделать эти правила эксплицитными и обязать кодировщиков соблюдать их в процессе кодирования.

Позднее этот подход был поставлен под сомнение Гарфинкелем², который доказал, что интерпретация полностью зависит от контекста. Марта Фельдман³ подытоживает позицию Гарфинкеля следующим образом: «Итак, понятие правила становится неадекватным, потому что для каждого контекста должно быть свое правило. Так как любой контекст уникален и постоянно возникают все новые и новые контексты, не может быть набора заранее разработанных правил, которые предполагалось бы соблюдать».

В теории можно конечно разработать очень сложный набор правил кодирования, которые будут определять, как элемент текста X должен быть закодирован, если он находится в контексте Y. Однако на практике количество возможных контекстов бесконечно. Поэтому такая стратегия вряд ли приемлема. Так как контекст оказывает особенно сильное влияние на коннотативные интерпретации, обычные правила кодирования не подходят для интерпретативного контент-анализа. Коннотативное кодирование в связи со сложностью содержания используемых категорий требует предельно высокого уровня экспертизы. Например, исторический анализ требует детального понимания исторического контекста, в котором создавались исследуемые сообщения. Публикации современных средств массовой информации также могут требовать достаточно глубоких знаний современных политических процессов, расстановки и взаимодействия политических сил в обществе. Поэтому лучше чтобы коннотативное

¹ Feldman, Marta S. *Strategies for Interpreting Qualitative Data*, Thousand Oaks, CA: Sage (1995).

² Garfinkel, H. *Studies in Ethnomethodology*, Cambridge, MA: Polity (1967)

³ Feldman, Martha S. *Strategies for Interpreting Qualitative Data*, Thousand Oaks, CA: Sage (1995)

кодирование или проводилось совместно несколькими кодировщиками, с постоянным обсуждением спорных моментов кодирования, или этим занимался один исследователь, обладающий теоретической чувствительностью. Теоретическая чувствительность является «персональным качеством исследователя, умением распознавать тонкости значений анализируемых данных».¹ Так как люди имеют различный уровень теоретической чувствительности, достижение согласия между кодировщиками при коннотативном кодировании весьма затруднительно и не всегда возможно.

Поскольку кодирование коннотативных значений избегает применения правил кодирования по вышеназванным причинам, очевидно, что имеется серьезное противоречие между потребностями коннотативного кодирования и традиционной контент-аналитической методологией. Поэтому интерпретативный контент-анализ предлагается как альтернатива традиционному контент-анализу для осуществления количественной обработки коннотативных интерпретаций.

Описание интерпретативного контент-анализа

Интерпретативный контент-анализ отличается от традиционного способом, которым проводится кодирование и тем, как оценивается качество кодирования. В традиционном контент-анализе несколько кодировщиков независимо друг от друга, используя правила кодирования, кодируют одни и те же тексты, что дает возможность охарактеризовать полученные с помощью такой процедуры данные как объективные, надежные и достоверные. При интерпретативном контент-анализе в принципе достаточно одного высококвалифицированного кодировщика, обладающего теоретической чувствительностью (обычно это сам автор проекта).

Традиционные контент-аналитики могут не подвергать кодированию анализируемый текст весь целиком. Вместо этого они могут сосредотачиваться на какой-либо его части, например, на заголовках статей. К сожалению, как пишет Ганс Кепплингер², в процессе кодирования «связная информация... разделяется на значимые части... При разделении связной информации на части взаимосвязями между частями обычно пренебрегают». Интерпретативный контент-анализ также может фокусироваться на частях текста (например, заголовке) как на единицах анализа. Но при кодировании исследователи руководствуются целостным подходом и учитывают то, как оставшиеся части текста влияют на интерпретацию кодируемой части. Современный взгляд на то, как люди приходят к пониманию текстов, подчеркивает то, что читатели интерпретируют любую часть текста, будь то слово или фраза, руководствуясь смыслом текста в целом. И в интерпретативном контент-анализе при кодировании частей текста последние не рассматриваются изолированно от остального текста, в котором они содержатся. А поскольку интерпретативный контент-анализ не ограничен правилами кодирования, он достаточно гибок для того, чтобы максимально полно и адекватно учитывать контекст.

В связи с тем, что в интерпретативном контент-анализе отсутствуют правила кодирования, при оценке его результатов неизбежно встает проблема качества кодирования, его объективности. Иными словами, как понимается объективность в традиционном контент-анализе, предполагается возможность получения тех же или максимально близких к ним результатов с помощью других кодировщиков. Но, к сожалению, в традиционном контент-анализе кодировщики, проинструктированные по поводу строгого следования правилам, пренебрегают своей интуицией. И тогда объективность результатов ни в коем случае не свидетельствует о том, что кодирование отражает популярную или широко распространенную интерпретацию данных текстов в среде их потребителей.

Интерпретации по сути своей не могут быть объективными. Любой исследователь осуществляет интерпретацию со своей собственной позиции, поэтому всегда найдется место для конкурирующих интерпретаций, сделанных с других позиций. В связи с этим можно утверждать, что для доказательства обоснованности своей точки зрения исследователю достаточно продемонстрировать при необходимости то, что его вариант кодирования, по крайней мере, в той же степе-

¹ Strauss, Anselm and Juliet Corbin, *Basics of Qualitative Research*, Newbury Park, CA: Sage (1990), p.41.

² Kepplinger, Hans M., *Content Analysis and Reception Analysis*, *American Behavioral Scientist*, 33 (1989), p. 177.

ни оправдан и правдоподобен, как и другие альтернативные варианты. Кроме того, исследователи часто не имеют равный уровень проникновения в природу интерпретируемых данных, и это тоже является причиной существования разных интерпретаций. Поэтому вполне правомерна точка зрения Холта¹, который пишет, что «об интерпретациях надо судить по их результативности... и их способности убедить читателя, не более того». Интерпретативный контент-анализ полностью принимает этот подход.

В некоторых случаях, чтобы продемонстрировать достоверность полученных данных исследователи предоставляют основание кодирования центральных текстов и сами эти тексты наряду с рукописью при передаче материалов для публикации. Таким образом, качество их кодирования может быть непосредственно оценено рецензентами. В случаях очень большого объема текстов могут быть представлены случайные образцы закодированных текстов. Если некоторые из кодирований не были прямыми, авторы могут предоставлять краткое обоснование их кодирования. Что это даёт? Прежде всего, это обеспечивает меру качественного контроля. В традиционном контент-анализе надёжность обеспечивается соблюдением правил кодирования, в интерпретативном же контент-анализе хотя и существует согласие между кодировщиками, но оно, как уже отмечалось, не обязательно отражает популярную или широко распространённую точку зрения на различные аспекты содержания кодируемых текстов, поэтому необходимо обращение к рецензентам. При этом может возникнуть проблема конкурирующих интерпретаций, поэтому роль рецензентов заключается в том, чтобы удостовериться в том, что предложенные интерпретации могут быть, безусловно оправданы имеющимися данными, даже если эти интерпретации не идентичны их собственным. В остальном же интерпретативный контент-анализ, подобно традиционному, является методом для количественного представления текстов в виде набора категорий при помощи кодирования.

Контент-анализ, основанный на ментальном восприятии или рецептивный

Кроме интерпретативного контент-анализа Ахувиа предлагает такую модификацию традиционного контент-анализа как контент-анализ основанный на ментальном восприятии или рецептивный. В чем его суть? Взгляд классика контент-анализа Бернарда Берельсона на эту проблему заключается в том, что контент-анализ ограничивается ситуациями, в которых виртуальный некто согласен со значениями анализируемых текстов. «Контент-анализ предполагает, что... содержание должно восприматься как «некая общая почва» для коммуникатора, аудитории и аналитика. Вот почему контент-аналитик предполагает, что «значения», которые он приписывает содержанию, распределяя их по определенным категориям, соответствует «значениям», которые имел в виду коммуникатор, и/или извлекает из текста аудитория»². Такое действительно может происходить в некоторых случаях, но как отмечает Кепплингер, «общая почва, насколько мы знаем это сегодня, возможно даже более ограничена, чем предполагал Берельсон»³. Джекоби, Хойер и Шелуга⁴ продемонстрировали в своих работах, насколько часто даже кажущиеся совершенно четкими и ясными утверждения по-разному понимаются разными людьми. И в таком случае встает вопрос: измеряем ли мы то, что предполагалось измерять, если центральной позицией в исследовании является проблема воздействия изучаемых текстов на определенные социальные группы (например, женщин, социальные меньшинства) Именно для решения этого вопроса и предназначен рецептивный контент-анализ, который является синтезом качественного рецептивного анализа и количественного контент-анализа.

Рецептивный анализ основывается на простом предположении: если вы хотите знать, как некая группа людей понимает набор текстов, имеет смысл спросить об этом у них самих. «Рецеп-

¹ Holt, Douglas B., *Rashomon Visits Consumer Behavior: An Interpretive Critique of Naturalistic Inquiry*, in *Advances in Consumer Research*, Vol.18, eds. R.Holman and M. Solomon, Provo, UT: Association for Consumer Research (1991), p. 16.

² Berelson, Bernard, *Content Analysis in Communication Research*, Glance (1952), p.19.

³ Kepplinger, Yans M., *Content Analysis and Reception Analysis*, *American Behavioral Scientist*, 33 (1989), p.175.

⁴ Jacoby, Jacob, Wayne, Hoyer, and David Sytluga, *Miscomprehension of Televised Communication*, New York: American Association of Advertising Agencies (1980). Hoyer, Wayne and JacobJacoby, *Miscomprehension of public Affairs Programming*, *Journal of Broadcasting and Electronic Media*, 29 (1985), p.p. 437-443.

тивное исследование» — это общий термин для исследований, в которых читателей спрашивают о том, как они понимают отдельные тексты. Этот качественный подход обычно детально изучает то, как читатели интерпретируют небольшое количество текстов, тогда как контент-анализ как правило практикует менее детальный разбор большого количества текстов. В названии предлагаемой модификации объединение терминов «рецептивный» и «контент-анализ» дает понять, что это некий «fusion» рецептивного исследования и контент-анализа, где сами читатели как центральные интерпретаторы выступают в роли кодировщиков текстов для перевода их в количественные показатели. Рецептивный контент-анализ сходен с традиционным контент-анализом в плане отбора анализируемых текстов, но отличается от него отбором, подготовкой и количеством кодировщиков, оценкой согласованности между ними, способом представления результатов... При этом кодировщики кодируют тексты не на основе жёстких чётких правил, а на основе своих интуитивных представлений.

В традиционном контент-анализе в качестве кодировщиков обычно выступают исследователи, или их студенты и аспиранты. В рецептивном контент-анализе кодировщики подбираются из той социальной группы, которая интересует исследователя в плане характера воздействия на нее анализируемых текстов. В традиционном контент-анализе кодировщики специально обучаются, для того чтобы четко следовать формальным правилам кодирования, поскольку «детально разработанные правила и процедуры уменьшают пристрастность субъективных суждений»¹. В рецептивном контент-анализе целью является не устранение субъективности, а напротив замеры особенностей субъективных восприятий текстов центральными интерпретаторами. Поэтому кодировщики осуществляют кодирование соответственно своему собственному интуитивному пониманию смысла текстов, а не следуя предписаниям правил.

В традиционном контент-анализе согласие между кодировщиками считается важным признаком качества, поскольку оно указывает на то, что кодировщики строго соблюдали правила кодирования. Но так как в рецептивном контент-анализе нет правил кодирования, согласие между кодировщиками не может рассматриваться как показатель качества процесса кодирования. В традиционном контент-анализе если две трети кодировщиков закодировали текст иначе, чем оставшаяся треть — это проблема. В рецептивном контент-анализе — это основание для научных выводов. В данном случае нет причины, обязывающей индивидуальных интерпретаторов соглашаться друг с другом по поводу распознавания текстовых значений. Любая попытка достичь консенсуса не естественным путем, не исходя из личной интуиции, повредила бы результатам исследования. Результаты рецептивного контент-анализа должны детально описывать все различия понимания текстов между кодировщиками, принадлежащими к разным социальным группам, в то время как традиционный контент-анализ приводит статистические данные, касающиеся выборки в целом. Рецептивный контент-анализ требует значительного большего числа кодировщиков по сравнению с традиционным, чтобы имелось достаточное пространство для статистического сравнительного анализа.

Предложенные варианты модификаций традиционного контент-анализа открывают новые возможности для исследователей. Контент-анализ, основанный на ментальном восприятии, позволяет составить более чёткое представление о воздействии текстов на общественное сознание различных социальных групп, даёт возможность исследовать различные способы, которыми одни и те же тексты могут быть поняты различными аудиториями. Интерпретативный контент-анализ поможет исследователям, тяготеющим к качественным методам исследования, примириться с использованием количественных методов традиционного контент-анализа. Также интерпретативный контент-анализ сможет облегчить изучение скрытых слоёв содержания и позволит применить более целостный подход к текстам средств массовой информации. Эти два подхода, признавая, что каждая интерпретация должна быть сделана с позиции специфической субъективной перспективы, а не с позиции объективной основы, открывают новые возможности для контент-анализа с применением более сложных интерпретаций анализируемых текстов.

¹ Kolbe, Richard H. and Melissa S. Burnett, Content-Analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity, *Journal of Consumer Research*, 18 (1991), p.245.

Литература и практики исключения: дискурс-анализ произведений, посвященных тюремной субкультуре

Тищенко Наталья Викторовна
*Саратовский государственный технический
университет*

Понятие дискурс на сегодняшний день является неотъемлемой частью многих теоретических построений и эмпирических исследований. Распространение дискурс-аналитических исследований объясняется отчасти универсальностью самого понятия «дискурс», которое охватывает всю совокупность текстов, производимых людьми в их повседневных практиках. Отчасти, интерес к дискурс-анализу вызван отсутствием ограничений: исследователь относительно свободен в выборе дизайна исследования, техники сбора эмпирического материала, способов оформления результатов и пр. Мы попытаемся на примере дискурс-анализа отечественных литературных произведений XX века, написанных непосредственно в местах лишения свободы или являющихся воспоминаниями авторов о пребывании в лагерях, колониях, тюрьмах, выявить основные теоретические предпосылки и установки дискурс-аналитического исследования. Цель работы заключается в раскрытии способов, стратегий, с помощью которых различные дискурсивные практики продвигают и утверждают определенные версии реальности¹; мы хотим выяснить, какой образ тюремной жизни формирует художественный, литературный дискурс. Дискурс-анализ является составной частью общего исследовательского проекта автора, посвященного изучению способов репрезентации тюремной субкультуры научным знанием, публицистикой, художественными произведениями и выявлению последствий этих репрезентаций для остальных социальных практик.

Теоретические основания. Анализируемый материал мы разделили на две группы: (1) литературные произведения, написанные профессиональными авторами (литераторами, журналистами, учеными); (2) произведения, написанные непрофессиональными авторами (это авторы, которые до того, как оказались в тюрьме, не занимались литературным творчеством, и именно опыт тюремного заключения побудил их созданию литературных текстов). Подобное разграничение эмпирического материала принципиально для дискурс-анализа, так как «профессиональные» и «непрофессиональные» авторы являются участниками двух различных дискурсивных конструкций, использующих различные знаковые, мифологические системы, иначе интерпретирующие базовые, ключевые понятия и, самое важное, противоположным образом трактующие конфликты, антагонизмы, противостояния внутри тюремной субкультуры. В исследовании учитывается и хронологическое разделение текстов, т.к. специфика отношения официальной власти к системе наказания и осужденным определяла и особенности ответной реакции в текстах, созданных в местах лишения свободы. Получилось следующее распределение эмпирического материала по группам:

1) профессиональные литераторы:

— *период репрессий первой половины XX века:* А.В. Жигулин «Черные камни», А.И. Солженицын «В круге первом», Д.С. Лихачев «Воспоминания», В.Т. Шаламов «Колымские рассказы», Л.Э. Разгон «Непридуманное», Е.С. Гинзбург «Крутой маршрут», Л.З. Копелев «Хранить вечно»;

— *советские и российские тюрьмы второй половины XX века:* Ф.Г. Светов «Тюрьма», Ю. Даниэль «Я всё сбиваюсь на литературу...», А. Марченко «Мои показания»;

2) непрофессиональные авторы:

— *период репрессий первой половины XX века:* М.П. Новиков «Воспоминания о пребывании в Бутырской тюрьме 1924–25 гг.»;

— *советские и российские тюрьмы второй половины XX века:* В.А. Ажиппо «Не зарекайся», И. Костов «И равнодушно смотрят небеса», Н. Лопатин «Записки зека», Л. Габышев

¹ Такую общую цель для любого дискурс-аналитического исследования формулирует Т.А. ван Дейк; см. Ван Дейк Т.А. Язык. Познание. Коммуникация. М.: Прогресс, 1989.

«Одлян или Воздух свободы», А. Павлов «Должно было быть не так», К. Гошев «Мужики в «Тишине». Опыт тюремного портрета».

Насколько оправдано использование литературных источников для дискурс-анализа? Главное отличие дискурса от литературного текста заключается в том, что дискурс представляет собой актуально произнесенный текст, имеющий отношение к конкретной ситуации, в которой он имеет значение. Другой существенной особенностью дискурса является то, что в нем не только отражены языковые формы высказываний, но содержатся оценочная информация, личностные и социальные характеристики участников дискурса. В дискурсе необходимо установить глубину связи описываемого события с индивидами, а так же определить степень вовлеченности участников дискурса в произносимый ими текст, т. е. определить степень уверенности индивидов в истинности их поступков, слов, убеждений. Следовательно, в процессе исследования необходимо ответить на вопрос: как дискурсы активируются анализируемыми текстами. Поэтому перечисленные произведения будут рассматриваться нами в качестве частей конкретных дискурсивных практик, содержащих в себе определенные мифологические структуры, способы идентификации, стратегии борьбы и гегемонии, а не как продукты индивидуального художественного творчества.

Дискурс-анализ предполагает мультиперспективное изучение объекта. В процессе исследования сопоставляются различные способы интерпретации того или иного явления и выявляется то новое знание об объекте, которое каждый из этих подходов предлагает. Дискурс-анализ дает исследователю возможность использовать несколько способов определения одних и тех же категорий, фактов, событий. Но этот метод предполагает и одно важное допущение — это единство теории, подхода и метода. Т.е. в основе любого дискурс-аналитического исследования лежит теоретическая база, включающая определенные интерпретации основных категорий — дискурс, дискурсивная практика, социальная практика, власть, идеология, гегемония, антагонизм и пр. Еще одним важнейшим допущением для дискурс-анализа является понимание сущности знания. Из данного подхода почти полностью элиминируется проблема истинности знания. Как пишет М. Фуко, «описание дискурсивных событий ставит вопрос: почему такие высказывания возникают именно здесь, а не где-либо ещё?»¹. Вопрос об истинности знания уступает место вопросу об особенностях формирования знания. Знания отличаются друг от друга не истинностью или ложностью, а формами организации. И различие формирования знаний определяется тем, где именно было сформулировано то или иное знание. Поэтому одной из принципиальнейших теоретических проблем в дискурс-анализе является проблема соотношения дискурсивных и социальных пространств. Ряд исследователей их отождествляют, другие же, напротив, разводят эти два вида пространств, но оговаривают целый ряд зависимостей между ними². Для нас этот вопрос так же является принципиальным, так как в исследовании речь пойдет о способах формирования знания в таком социальном пространстве как пенитенциарная система. С одной стороны изолированность этого социального пространства от других социальных систем, с другой — принудительное помещение индивида в это пространство определяют особенность тех дискурсивных практик, которые формируются внутри тюремной системы и соперничают между собой за право предложить наиболее адекватное и полное описание того социального порядка, который мы будем называть тюремной субкультурой. Важно отметить, что существуют ещё и внешние пенитенциарной системе дискурсы, которые так или иначе интерпретируют такое явление как тюрьма и весь связанный с ней порядок высказываний (преступление, заключенный, право, система наказания и т.д.). Есть дискурсы тюремной субкультуры и есть дискурсы о тюрьме: либеральный («модифицировать тюрьмы». «сокращение наказаний, связанных с лишением свободы»), консервативный («ужесточение мер наказания»), тоталитарный («отмена моратория на смертную казнь»). Это так называемые общегражданские или политические дискурсы. Есть так же дискурсы о тюрьме профессиональные: криминологи-

¹ Фуко М. Археология знания. Киев: Ника-Центр, 1996. С. 29.

² Fairclough N. Critical discourse analysis and the marketization of public discourse: die universities // Discourse and Society. 1993. № 4 (2). P. 133–168.

ческий, судебный, медицинский, правозащитный (права осужденных, статус политзаключенных), социология преступления, социология тюремного заключения и пр. Мы исследуем дискурсы, которые формируются в пространстве тюрьмы, но они, несомненно, имеют определённые связи и пересечения с внешними дискурсами о тюрьме.

Одно из первых определение дискурса было предложено М. Фуко, который под дискурсом понимал формы закономерности и специфику организации дискурсивных событий. Основным элементом дискурсивных событий, согласно Фуко, являются высказывания, которые не исчерпываются лишь языковыми построениями, а включают мир материальных объектов, правила действия, принципы мышления, производимые в определенную эпоху в определенном социальном пространстве¹. Принципиальной для Фуко является властная функция дискурса, благодаря которой он объясняет каким именно образом смыслы и значения, разделяемые изначально лишь одной профессиональной, социальной группой, приобретают со временем статус общепризнанного знания, научной дисциплины, истины. Развитие властной функции дискурса Фуко разделяет на четыре стадии, четыре уровня конструирования дискурса, которые он определяет как «пороги»:

- 1) порог позитивности: дискурсивные практики достигают индивидуализации и автономии;
- 2) порог эпистемологизации: дискурсивные практики осуществляют доминирующую функцию в отношении знания;
- 3) порог научности: выработка формальных критериев;
- 4) порог формализации: возникновение научной дисциплины².

Преодолев все четыре порога, дискурс становится неотъемлемой частью всех социальных практик, используемых в данной культурной среде, до тех пор, пока иной тип дискурса не сумеет предложить более актуальные способы объяснения и не займет лидирующих позиций среди социальных практик. Тюрьма (в современном понимании исправительная, пенитенциарная система) в качестве своеобразного социального института является объектом рассуждения многих официальных, легализованных дискурсов. Каждый из этих дискурсов имеет свою историю возникновения и развития. Объединяет же их всех происхождение; своим появлением они обязаны (1) развитию широкого либерального дискурса, соотносящегося с конкретными социальными практиками (либеральной политикой, рыночной экономикой), и (2) развитию корпуса наук о человеке, в котором человек представлен во множествах образах, в том числе и как человек-преступник. Особенностью этих официальных дискурсов является то, что, пройдя все пороги конструирования и заняв свою дисциплинарную нишу, ни один из этих дискурсов не признается самими представителями тюремной субкультуры, т.е. осужденными. Мы не станем распространять наш вывод на зарубежные пенитенциарные системы, но отечественная тюремная субкультура является ярким примером противостояния и противодействия любому официальному дискурсу о тюрьме.

Далеко не все современные авторы, применяющие дискурс-анализ, рассуждают в контексте властных отношений, борьбы дискурсов и пр. Для многих авторов различные дискурсы, связанные с функционированием социальных институтов, вполне мирно уживаются друг с другом, создавая плюралистичное, контрастное поле смыслов и значений³. Но в данном исследовании основное внимание уделяется именно властной функции дискурса. На наш взгляд, борьба различных дискурсов за право дать наиболее точную, «правдивую» интерпретацию тюрьмы, тюремной субкультуры, тюремной жизни определяет специфику социальных практик, которые складываются относительно такого социального пространства как пенитенциарная система. Показательно, что идеи *правдивости*, *откровенности* используются как официальными дискурсами, так и дискурсами принципиально отрицающими любые официально признанные формы высказывания о тюрьме. Показательной здесь может быть цитата из редакторского вступления к книге Л. Разгона «Непридуманное» (хотя уже само на-

¹ Фуко М. Археология знания. Киев: Ника-Центр, 1996. С. 28–30.

² Там же. С. 185–186.

³ Kress G., van Leeuwen T. Multi-Modal Discourse: The Modes and Media of Contemporary Communication. London: Arnold, 2001.

звание текста четко указывает на отношение автора к своему тексту): «Все, о чем рассказывает в этой книге писатель Лев Разгон, — правда. В ее обычном словарном обозначении: «То, что действительно было, то, что в действительности есть». В ней нет придуманных персонажей, эпизодов, дат»¹. Столкновения и взаимное отрицание официальных и неофициальных дискурсов о тюрьме является ярчайшим примером того, каким образом в обществе формируются и транслируются идеи, мифы, легенды о различных объектах, как возникают образы «героев» и «антигероев» в зависимости от того, какой из дискурсов оказывается наиболее чувствителен и адекватен действующим социальным практикам.

Еще одним важным концептом в дискурс-анализе является понятие «идеология». Ключевыми теоретиками понятия идеологии в дискурс-аналитических исследованиях многие авторы называют Л. Альтюссера и А. Грамши. Альтюссер считал, что сопротивление идеологии невозможно, т.к. идеология — это искажение реальности и индивид просто помещен в эту искусственную созданную идеологическими конструктами реальность². Грамши полагал, что гегемония — это социальное согласие, подменяющее групповые (классовые) интересы, и индивиды в любой идеологической ситуации способны артикулировать свои интересы и приспособлять к ним идеологический строй³. Последующие авторы, занимающиеся дискурс-анализом, полагают, что Альтюссер недооценивал способность людей противостоять и сопротивляться идеологии, конструируя собственные дискурсивные практики. На наш взгляд, литературные произведения, написанные осужденными, являются ярким примером подобного сопротивления идеологии, которая выражена в таких социальных практиках как судебная система, тюремный режим и пр. Чем жестче идеология и агрессивней социальные практики, актуализирующие её, тем больше появляется примеров конструирования ответных дискурсов, в нашем случае, текстов, рассказывающих о пребывании в местах лишения свободы.

Идеологию можно интерпретировать в качестве «оклика» власти, на который гражданское общество всегда вынуждено реагировать с той или иной быстротой и интенсивностью. В тюрьме этот оклик максимально агрессивен, безапелляционен, он должен получать моментальный ответ от осужденных, причем ответ, заранее предусмотренный и предписанный властью. «Оклик» власти всегда ограничивающий и дискриминирующий. Тогда литературные произведения, написанные в тюрьме — это игнорирование оклика или неадекватный этому оклику ответ. С точки зрения официального дискурса, это неприемлемая норма поведения. Но и с точки зрения тюремной субкультуры, литература о тюрьме так же является неким нонсенсом, поскольку авторы, как правило, критически относятся не только к официальным, режимным правилам в местах лишения свободы, но и осуждают нормы и принципы уголовного мира.

Нужны ли подобные тексты тюремной субкультуре и дискурсу, который ей соответствует? Исследуя образы тюремной субкультуры в кинотекстах, мы столкнулись с любопытным примером восприятия кинофильма о тюремной жизни представителем Управления наказания и самих осужденных. Вот цитата из интервью начальника Управления ФСИН по Тверской области А.А. Савихина на фильм «Беспредел»: «К сожалению, я не могу вспомнить по настоящему правдивого фильма о колонии. Например, в нашей 10-й колонии снимался фильм «Беспредел», где допущено множество неточностей. Например, начальник оперчасти, которого играл Лев Дуров, провоцирует конфликты среди осужденных, хотя это по определению не в его интересах. Такого начальника уволили бы за подобные игры в два счета. Первыми зрителями фильма стали осужденные из 10-ки. Они прямо сказали режиссеру Гостеву, что фильм неправильный, но он им заявил, что, мол, вы свою жизнь не знаете, а я знаю. Ну что здесь можно комментировать. Про фильм «Зона» вообще не стоит говорить. В нем персонажи показаны так, как нужно авторам, которые хотят пугать и шокировать публику, а не так, как все есть в жизни»⁴. Самыми значимыми для анализа здесь являются слова режиссера в пересказе начальника УИН: «*вы свою жизнь не знаете, а я знаю*». Только взгляд извне,

¹ Разгон Л.Э. Непридуманное. М.: Изд-во «Захаров», 2007. С. 4.

² Альтюссер Л. Ленин и философия. М.: Ad Marginen, 2004. С. 39–40.

³ Грамши А. Тюремные тетради. М.: Политиздат, 1991. С. 85.

⁴ Гражданин начальник 09.08.2006. // Режим доступа: <http://www.etver.ru>

взгляд исследователя, не обремененного субъективными пристрастиями и переживаниями, профессиональными стереотипами, способен показать и рассказать истину объекта изучения, в данном случае тюремную жизнь. Однако те, кто непосредственно находится внутри мира тюрьмы, в корне не согласны с объяснением внешнего наблюдателя. Их взгляд на события на зоне, на действия руководства колонией, на правила и установки тюремной жизни, на криминальную иерархию диаметрально противоположны взгляду исследователя-режиссера.

Ещё один пример отрицания правоты взгляда наблюдателя за тюремной жизнью встречается в тексте В. Ажиппо. Этот автор, прослужив несколько десятилетий в исправительной системе, создает текст, который противопоставляет всем прочим способам интерпретации тюремной жизни: «То, что вы увидите в первый момент в камере, никак не будет походить на дурацкие картины из дешевых книжек и кинофильмов. Никто на вас не станет рычать, не будет татуированных амбалов, которые сразу же попробуют вынуть из вашего рта золотой зуб. Все это нездоровые фантазии литераторов и киношников, рассчитанные на такой же болезненный интерес к тюрьме со стороны обывателя»¹.

С точки зрения самих авторов литературных произведений, написанных в местах лишения свободы, их тексты — это, во-первых, попытка создать своё индивидуальное, приватное пространство, которое невозможно с физической точки зрения в местах лишения свободы, во-вторых, конструирование такого типа дискурса, который одновременно компрометирует и раскрывает все недостатки и дискурса власти, и дискурса тюремного мира, но и формирует подлинный образ системы изоляции. Что очень характерно, кем бы ни был автор текста о тюрьме — политзаключенный, уголовный преступник, сотрудник системы наказания — он противопоставляет свой текст всем остальным системам высказывания и обещает читателю, что тот будет посвящен в самое сокровенное, что всегда всеми замалчивалось и что составляет сердцевину не только тюремной субкультуры, но и человеческого существования вообще. Очень показательной здесь является фраза из предисловия А. Абросимова к работе В. Ажиппо «Не зарекайся»: «Если разобраться, то его книга *не о тюрьме, а о жизни*. Я надеюсь, что она многим поможет уверенно ориентироваться в окружающем мире, научиться видеть добро, без страха противостоять злу, беречь честь и хранить достоинство»² (*курсив. — Н.В.*). Неофициальный дискурс трактует тюрьму как особое социальное пространство, индивид получает доступ к пониманию экзистенциальных истин, которые в принципе недоступны в рутинном течении повседневной жизни в других социальных пространствах. Интерпретация тюрьмы как пространства постижения истины характерна в художественных текстах, как для представителей «прозы камерного жанра», так и для профессиональных авторов. Знаменитая фраза А.И. Солженицына «Благословение тебе, тюрьма, что ты была в моей жизни! А из могил мне отвечают — Хорошо тебе говорить, когда ты жив остался!»³ утверждает за тюрьмой статус пространства, в котором проблема выживания рассматривается в нескольких ипостасях — физическом, нравственном, сакральном. Независимо от того, считает ли автор себя виновным и заслуживающим наказание, тюрьма воспринимается им в качестве пространства, находящегося на грани социального и над-, сверхсоциального. Тюрьма — это пограничная область, и её статус вызывает некий трепет, уважение и искреннюю благодарность, как ни парадоксально это звучит. В другой книге А.И. Солженицына фраза, вложенная в уста персонажа Глеба Нержина, совпадает с предыдущей цитатой: «Благословение тюрьме!! Она дала мне задуматься. Чтобы понять природу счастья»⁴. Столь настойчивое повторение одного и того же оборота лишь подтверждает, насколько важен этот пограничный, сакральный статус тюрьмы. И ещё одна цитата, созвучная приведенным высказываниям уже из текста Ф.Г. Светова: «И таким я чувствовал себя свободным, богатым, так бывал счастлив, оставаясь один, забравшись головой в матрасовку, — один на один с Богом, с Его свя-

¹ Ажиппо В.А. Не зарекайся. Харьков: Батьківщина, 2005. С. 9.

² Абросимов А. Предисловие // Ажиппо В. Не зарекайся. Харьков: Батьківщина, 2005. С. 4.

³ Солженицын А.И. Архипелаг ГУЛАГ (ч. 4, гл. 1). Кемерово: Кемеровское книжное изд-во, 1990. Т. 2. С. 571.

⁴ Солженицын А.И. В круге первом. М.: АСТ; Астрель, 2007. С. 46.

тыми молитвами! Вчитайтесь, подумайте»¹. В результате тюрьма оказывается пространством, где подводятся итоги всего предыдущего существования, где раскрывается смысл всех поступков, мыслей, чувств индивида: «Каждый попавший в тюрьму, как правило, оказывается на том месте, которое он заслужил всей своей прежней жизнью»².

Инструменты эмпирического анализа. Методики проведения дискурс-анализа достаточно разнообразны. В качестве примера дискурс-аналитической методики можно привести схему, предложенную Р. Пирсом и состоящую из семи пунктов: (1) описание; (2) исследование коннотаций; (3) идентификация дискурса; (4) определение стратегий субъективации; (5) поиск похожих дискурсов в других текстах; (6) оценка исторических измерений; (7) выводы. Автор исходит из позиции, что нет «правильного» способа интерпретации текста, и с осторожностью относится к возможности точного/истинного прочтения текста³. Но существует и ряд направлений в дискурс-аналитических исследованиях, которые не предлагают четких определённых указаний, как именно проводить эмпирическое исследование, ограничиваясь определением теоретических оснований анализа. Так Э. Лакло и Ш. Муфф, исходя из принципа, что разнообразные дискурсы находятся в состоянии постоянной борьбы между собой, выделяют лишь несколько концептов, позволяющих проводить конкретное исследование: (1) основные знаки для определенного дискурса; (2) понятия, расшифровывающие основные знаки дискурса; (3) системы идентификации (способы самоидентификации и способы соотношения индивида с группой); (4) анализ существующих дискурсивных конфликтов⁴. За основу в исследовании произведений, посвященных тюрьме, мы взяли схему, предложенную в работах Э. Лакло и Ш. Муфф. Главным преимуществом этой схемы является то, что с её помощью можно восстановить всю цепочку дискурсивных конфликтов, которые возникают при конструировании такого явления как «тюрьма».

Основные знаки дискурса включают мифы, героев и антигероев, устойчивые выражения, характерные определения, используемые в той или иной дискурсивной конструкции. Основные знаки — это максимально общие, своего рода метафизические понятия. Это пустые знаки, они сами по себе ничего не значат, некие само собой разумеющиеся данности, не требующие расшифровки для участников одного дискурса и абсолютно непонятные, закрытые для интерпретации конструкции участниками другого дискурса. Поэтому и произведения профессиональных авторов, и в тексты авторов-любителей используют в большинстве случаев одни и те же знаки, которые принципиально отличаются цепочками эквивалентностей, с чьей помощью раскрывается смысл основных понятий. Особенностью дискурсов, сформированных внутри пространства тюрьмы, является то, что основные знаки и мифы этих дискурсов либо полностью игнорируются, либо принципиально иначе интерпретируются иными (официальными) дискурсами. Но и дискурсы, представленные в литературных текстах, не признают официальные определения и конструкции, явно иронизируя по поводу властной терминологии или критикуя её. Замечание из текста В.А. Ажиппо относительно названий лагерной системы коррелирует с отрывком из книги «В круге первом» А.И. Солженицына:

«Самое забавное, что неформальной лексике совершенно наплевать на изыски чиновников... тюремные определения гораздо точнее официальной терминологии»⁵;

«Озерлаг, Луглаг, Степлаг, Камышлаг...

— Можно подумать, в МВД сидит непризнанный поэт. На поэму не разгонится, на стихотворение не соберется, так дает поэтические названия лагерям»⁶.

¹ Светов Ф.Г. Тюрьма. [Электронный ресурс]. Режим доступа: Центр содействия реформе уголовного правосудия <http://www.prison.org>

² Абрамкин В.Ф., Чижов Ю.В. и др. Как выжить в советской тюрьме. В помощь узнику. Красноярск: Восток, 1992. С. 109.

³ Pearce R. Advertising: critical analysis of images // Critical textwork: an introduction to varieties of discourse and analysis. Buckingham: Open University Press, 1999. P. 78–91.

⁴ Laclau E., Mouffe C. Hegemony and Socialist Strategy. Toward a Radical Democratic Politics. London: Verso, 2001. P. 97–98.

⁵ Ажиппо В.А. Не зарекайся. Харьков: Батьківщина, 2005. С. 28.

⁶ Солженицын А.И. В круге первом. М.: АСТ; Астрель, 2007. С. 12.

Ещё одной специфической чертой, отличающей основные знаки, которые организуют дискурсы в местах лишения свободы, является их дихотомичность. Они представлены парой противоположностей, причем обе части узлового знака не отрицают друг друга, а составляют онтологическую общность: «зона – свобода», «зеки – вольные», «власть – беспредел», «порядок – насилие». Использование только лишь одной части основного знака бессмысленно в пространстве данных дискурсов. Дихотомичность основных понятий усиливает конфликтность анализируемых дискурсивных структур. Дискурсы, сформировавшиеся в местах лишения свободы, не только конфликтуют между собой и официальными дискурсами о тюрьме, но, самое важное, конфликтность располагается внутри самого дискурса. Именно поэтому ключевые высказывания, организующие тексты, могут противоречить друг другу или даже исключать друг друга:

- «тюрьма — это школа жизни» – «злой и подлый тюремный мир»;
- «только в тюрьме человек чувствует свободным» – «да, не свободной республикой оказался и лагерь»;
- «в тюрьме каждый сам за себя» – «в одиночку не выжить в тюрьме» – «тюрьма стирает границы между людьми».

Основные знаки, организующие дискурсы, определяют и характер мифов, которые конструируют социальное пространство тюрьмы, а так же черты героев и антигероев, действующих в этом социальном пространстве. В текстах писателей-любителей мифы связаны с традициями и правилами тюремной субкультуры, определяющими правила поведения тех, кто оказался в местах лишения свободы. Главная идея, организующая эти мифы, заключается в противопоставлении жестоких, но «справедливых» правил тюремной субкультуры, помогающих выжить в нечеловеческих условиях, и жестоких и «несправедливых» правил режима, санкционируемых администрацией тюрем: «Да, конечно, люди, отбывающие в местах лишения свободы определенные судом срока за различные преступления, виновны. И, конечно же, условия их содержания не должны быть похожи на отдых в санатории. Но не до такой же степени! А почему так должны страдать те, кто еще не признан судом виновным и находится под следствием. С таким отношением государства к своему народу нам никогда не стать нормальной демократической страной, никогда...»¹. Структура этих текстов достаточно похожа и определяется хронологией событий, которые происходят с индивидом, попавшим в места лишения свободы. В начале идут мифы, связанные с арестом и пребыванием в СИЗО, затем подробный рассказ о ритуалах и традициях камерной жизни с подробным объяснением тюремной иерархии как официальной, так и неофициальной. Ключевым лейтмотивом подобных текстов являются многочисленные предостережения, табу и запреты. Одним из главных советов является запрет на откровенные разговоры, как с сокамерниками, так и с администрацией. Утверждается принцип абсолютного, онтологического недоверия, которое пронизывает все сферы тюремной жизни. И следующий из этого запрета принцип: осторожное отношение к словам. Каждое слово может быть интерпретировано против говорящего, слова превращаются в амбивалентные, постоянно меняющие свои смыслы структуры. Поэтому молчание — это идеальная форма существования в местах лишения свободы: «Это вот странная такая игра в слова — все принимают всерьез»².

В произведениях профессиональных писателей присутствует совсем иная структура мифов, которая, прежде всего, связана с идеей испытания, поиска истины и обретения себя через потерю социального статуса и материальных благ. В отличие от авторов-любителей, которые многие страницы тратят на советы как сохранить продукты, как наладить отношения с администрацией и уголовным миром и пр., профессиональные авторы с большой долей иронии описывают лагерный быт, еду, одежду зеков, условия содержания. Внешние ограничения и потери — это лишь повод для шутки, пусть и злой иронии. Страшнее, когда индивид теряет внутренние, экзистенциальные ориентиры, и мифология произведений, написанных профессиональными авторами, связывается с попытками сохранить эти внутренние устои:

¹ Костров И. И равнодушно смотрят небеса. Режим доступа: <http://www.tyurem.net/books/kostrov/index.htm>

² Абрамкин В.Ф., Чижев Ю.В. и др. Как выжить в советской тюрьме. В помощь узнику. Красноярск: Восток, 1992. С. 111.

«Ими владело бесстрашие людей, утеравших все до конца, — бесстрашие, достающееся трудно, но прочно»¹.

Исследователи, применяющие дискурс-анализ, так или иначе корректируют методику соответственно анализируемому случаю, акцентируя внимание на нужном фрагменте. Пункт, который относительно стабильно остается неизменным в любых исследованиях, — соотношение текста с более широким историческим и социальным контекстом. Принципиальным для дискурс-анализа является ответ на вопрос, что значит тот или иной основной знак не только в рамках данного текста, но и в рамках данного общества в тот момент времени, когда текст появился.

В любом социальном пространстве есть разделяемая подавляющим большинством акторов цель — достижение профессионального успеха, материального благополучия и пр. Каждый дискурс формирует основные критерии, показатели этой цели. В социально одобряемых пространствах, в которых используются культурно приемлемые дискурсивные практики, положение акторов определяется их положением относительно всеми принимаемой цели. Система социальных статусов — это своеобразное графическое распределение акторов относительно основной цели. Те, кто находится наиболее близко к ней, те обладают большими социальными, культурными, символическими ресурсами. И связь индивидов в одном дискурсивном пространстве конфликтна в функциональном отношении. Особенностью такого социального пространства как тюрьма является отсутствие общей, связующей цели. И объединение между участниками дискурсивного пространства не строится на основе конфликтной цели. Конфликт в отношении тюрьмы носит онтологический характер. Участники дискурсивных практик, имеющих распространение в тюрьме, сами по себе конфликтны по отношению к любому легальному дискурсу: по отношению к тоталитарной идеологии конфликтны политические заключенные, по отношению к либеральным ценностям (жизнь, свобода, имущество) — уголовные заключенные. Поэтому участники дискурсивных тюремных практик не соперничают друг с другом в достижении общей цели (например, освобождению из тюрьмы, ведь у каждого свой определенный срок в зависимости от статьи наказания), этой общей, разделяемой цели в местах лишения свободы просто не существует. Это одна из причин, почему институт тюрьмы абсолютно неэффективный социальный институт и почему общества, организованные по принципу изоляции так же неэффективны, антипроизводительны. Борьба за символический социальный ряд социальных статусов заменяется в местах лишения свободы борьбой за физическое пространство, за место в камере, на нарах и пр. Основные термины тюремных дискурсов — отстраненность, недоверие, молчание, сам по себе и пр. — выражают суть главного противостояния в местах лишения свободы: конфликта между человеком и изоляцией. В текстах профессиональных писателей это конфликт выражается в экзистенциальном поиске самого себя и конечном обретении подлинного понимания своих эмоциональных, нравственных, ну и физических возможностей. Поэтому Солженицын и напишет: «Страшно подумать, что б я был за писатель, если б меня тогда не посадили»². Авторы же непрофессиональные, представители прозы камерного жанра ключевой конфликт интерпретируют в бытовом плане и эти тексты представляют собой памятку, руководство, курс по выживанию в социальном пространстве, где индивид всегда на виду, но всегда абсолютно одинок.

С точки зрения дискурс-анализа общество не может существовать как объективная данность: оно никогда не завершено благодаря дискурсивным практикам, всегда открытым для трансформаций³. В местах лишения свободы искусственно поддерживается «завершённый», объективный проект общества, там преследуется любое развитие дискурсов. И литература, независимо профессиональная или любительская, — это способ сопротивления законсервированности, закрытости дискурсивных практик в тюрьме.

¹ Солженицын А.И. В круге первом. М.: АСТ; Астрель, 2007. С. 426.

² Солженицын А.И. Бодался теленок с дубом. Очерки литературной жизни // Режим доступа: <http://biblioteka.cc/topic/35756-bodalsja-telenok-s-dubom-solzhenicin-aleksandr/>

³ Laclau E., Mouffe C. Hegemony and Socialist Strategy. Toward a Radical Democratic Politics. London: Verso, 2001. P. 113.

Этноавтобиография как методика анализа этнической идентичности

Уразалиева Гульшат,
Беккерова Елена Валерьевна,
РГГУ

Термин «этноавтобиография» был сконструирован для выполнения учебной задачи на 3 курсе социологического факультета РГГУ по дисциплине «Этносоциология». Эта учебная задача была направлена на то, что бы студенты смогли поразмышлять над основаниями своей этнической принадлежности. Им была предложена домашняя письменная работа, где они могли в свободной форме и в виде самостоятельного эссе изложить свои мысли. Никаких шаблонов выполнения не задавалось. Прежде чем эту задачу поставить перед студентами, я испробовала этот подход на себе, в виде своеобразного кейс-стади.

На первой лекции в начале курса рассказала студентам о себе, что значит для меня быть казашкой. Я им показала генеалогическое древо казахов, сделанное историками АН РК, на котором был изображен весь жузово-родовой строй казахов. Впервые о своей подробной родословной я услышала от своего деда Оразалы, который мне 9-класснице продиктовал 18 имен наших прадедов. Начал он с имени моего отца Кулумжана Уразалиева и закончил двенадцатью отцами Байлы, первопредками, основателями младшего жуза, один из которых — Кадыркожа — и стал предком нашей ветви. К моему великому удивлению, именно эта семейная родословная, составленная дедушкой, нашла позже подтверждение в Шежире (генеалогия) казахов, опубликованном историками Казахстана после перестройки. Этот мой рассказ впечатлил моих студентов социологического факультета, слушавших впервые по учебному плану курс «Этносоциологии». На следующий год я повторила свой экскурс в казахскую историю через семейную траекторию, однако позднее такого подробного рассказа о своей казахской идентичности не повторялось. Это было связано с развитием курса, включением новых тем.

Но именно этот опыт саморефлексии над основами своей этнической идентичности помог мне выйти с предложением написать студентам нечто подобное. Сформулировано задание впервые было в виде неоконченного предложения «Я русская (казашка, татарин и т.д.) — это значит...». Теоретической предпосылкой для меня послужила работа Ч. Миллса «Социологическое воображение», которое натолкнуло на мысль разбудить такое социологическое воображение у студентов. На кафедре в это время работал профессор Агдас Кусаинович Бурганов, который написал своеобразную «Социобиографию»; знакомство с его рукописью тоже помогло в кристаллизации будущей методики исследования идентичности через автобиографический текст.

Сбор ежегодных студенческих эссе, заданных как домашняя работа, позволил накопить уникальный корпус текстов, документов этнической идентификации молодого поколения начала XXI века. За 10 лет преподавания «Этносоциологии» все студенты писали такие тексты. Это стало фирменным знаком моего курса.

Три года прошло с начала сбора текстов этноавтобиографий, когда они стали основой самостоятельного направления по анализу конструирования этнической самоидентификации через написания текста. Постепенно учебная задача трансформировалась в учебную методику, применяемую мной уже десять лет для анализа этнической идентичности студентов и используется по сей день на СФ РГГУ. Эта методика была апробирована и на казахстанских студентах в 3-х вузах в городах Алматы и Тараз. Этот опыт работы вне российского поля показал, что методика работает в том же режиме, что и с российскими студентами. В результате описание термина приобрело следующую форму.

Этноавтобиография — методика анализа формирования этнической идентичности, представляющая собой письменный ответ в свободной форме на вопрос: «Я русский (татарин, еврей и т.д.), что это для меня значит?»

На вопрос «Я казах (русский, татарин, украинец, еврей, чеченец и т.д.), что это для меня значит?» казахстанские студенты отвечали тоже в совершенно произвольной форме, без подсказок и оценивания содержания со стороны моих друзей, казахстанских преподавателей.

Они мне пересылали оказией тексты в Москву. Ответы казахстанских студентов были, по сути, похожи на ответы российских студенты. Но там были и оригинальные ответы, которых не могло быть у россиян. Именно в этом проявилась удивительная жизнеспособность методики, что в массиве текстов казахских студентов появилось то, что не могло быть у русских казахстанцев, тем более россиян. Они написали о своих жузово-родовых корнях. Из всех текстов где-то десятая часть работ имела отсылки к своим родоплеменным корням. Именно эти тексты доказали, что произвольность в написании этноавтобиографии позволяет выйти на качественную методологию.

Методические требования к написанию этноавтобиографии заключались в первую очередь в соблюдении временных рамок. Было желание получить результат сразу после первой лекции по курсу «Этносоциология», чтобы последующая работа на лекциях и семинарах не перегружала студентов излишним теоретизированием. Этому способствовала система выставления баллов за домашнюю работу. Высшая отметка в 10 баллов ставится только за сам факт сдачи домашней работы в срок, то есть к следующей второй лекции. За опоздание на неделю снижается оценка на 1 балл. Никаких комментариев, как именно надо написать работу мной не давалось. В аудитории на семинарах содержание работ не комментировалось. Студенты были уверены, что искренность им не повредит. Через год я обобщила первый раз итоги домашних заданий и использовала в научной статье. Студенты участвовали в обработке и впервые результаты были обнародованы в 2002 году в статье А. Дубининой и доложены на научной студенческой конференции. Позже я предложила писать по текстам этноавтобиографий курсовую и дипломную работу. Елена Беккерова написала и защитила диплом на тему «Особенности русской идентичности в этноавтобиографиях студентов» в 2009 году.

Для исследования конструирования русской идентичности нами были отобраны тексты этноавтобиографий 180 студенческих эссе, написанных в 2006–2009 годы. Респондентами выступили студенты 3 курса социологического факультета РГГУ различных форм обучения.

Использовался метод дискурсного анализа, который позволяет выявить, каким образом выбор студентами тех или иных риторических и языковых средств определяет способы конструирования события, сюжета и повествования, в нашем случае, их этнической идентичности. Кроме того, был применен контент-анализ для выявления смысловых единиц, используемых респондентами в этноавтобиографиях.

Тексты этноавтобиографий стали единицами анализа, на основе которых и производился подсчет частоты встречаемости категорий. Категории контент-анализа объединяют множество разрозненных слов-признаков в тематическое, концептуальное целое и обозначают эмпирические признаки текстовой информации, сформированные в результате операционализации основных теоретических понятий исследования. Поскольку в дипломной работе был рекомендован структуралистско-конструктивистский подход к изучению этнической идентичности, при проведении контент-анализа за категории были приняты, в первую очередь, механизмы конструирования этноса, изложенные в работе В. Ильина¹. В связи с этим, были выделены следующие категории, каждая из которых идентифицировалась по набору словоформ (см. Таблица 1).

Изначально предполагалось, что осознание своей принадлежности к тому или иному этносу приведет также и к осознанию времени и условий, при которых происходила идентификация. Однако анализ работ показал, что далеко не все студенты затрагивали вопрос формирования своей этнической идентичности. В большинстве этноавтобиографий говорилось о значении своего этноса, его составляющих, о причинах причисления себя к той или иной этнической группе.

Как уже отмечалось, одним из самых важных механизмов конструирования этнической идентичности является этническое самоназвание. Так, из 180 опрошенных студентов 169 идентифицировали себя с русским этносом, причем 13 из них дети межэтнических браков, выбравшие для себя русскую идентичность.

¹ Ильин В.И. Социальное неравенство. М.: Институт социологии РАН, 2000.

Категории контент-анализа

Самоидентификация	Я русский; не могу сказать; не могу определить; затрудняюсь
Идея общности происхождения и судьбы	Корни; история; общий
Этническая принадлежность членов семьи	Родители; дедушка; бабушка; корни; предки
Менталитет	Менталитет; характер; русская душа; черты; состояние души
Язык	Язык; родной
Внешняя идентификация	Внешность; выгляжу; волосы; глаза
Стереотипы	Стереотипы
Религия	Вера; православие; православный; христианин; христианство; религия; обряды
Культура	Культура; обычаи; традиции; праздники; пасха; рождество
Национальное государство	Родился; живу; Россия; россиянин; страна; государство; родина
Этнические чувства	Гордость; горжусь; стыд; стыжусь; рад; счастлив; люблю; грущу

Один из студентов, хотя и определил себя как русского, однако, склонялся к более широкому понятию «славяне»: *«Я не могу назвать русский этнос именно этносом, скорее нацией. Мы — славяне, принадлежащие к славянскому этносу. Именно восточнославянский этнос ввел под один многие этносы и национальности и ввел единый термин, единое определение»* [2, 145¹]. Эта своеобразная надэтническая идентичность славян в ходу у многих русских людей, что подтвердилось и в текстах студентов. У остальных 11 человек родители также принадлежат к различным этническим группам, в связи с чем, их дети не смогли себя идентифицировать лишь с одной из них, говоря о своей двойной идентичности. В процессе анализа были выделены категории, по которым респонденты причисляли себя к тому или иному этносу. Большинство из них совпали с механизмами конструирования этнической общности (самоидентификация, язык, внешняя идентификация, стереотипы, религиозная принадлежность, идея общности судьбы, народные обычаи, официальная идентификация).

Однако результаты исследования отличаются от результатов, полученных другими авторами. Оказалось, что половина студентов идентифицируют себя с русским этносом, в первую очередь, потому, что они воспитывались в русской культурной среде, им близки ее обычаи и традиции, они любят и знают творчество русских писателей, музыкантов и художников. Именно воспитание в русской культуре стало для некоторых детей межэтнических браков причиной идентификации с русским этносом. Как отмечает В.И. Ильин, следование обычаям превращает этнические границы в видимые, легко распознаваемые. *«Я — русский, я один из тех, кто соблюдает, сохраняет и должен передать новому поколению культуру, нравы, традиции, обычаи, которые присущи нашей народности»*, — пишет один из респондентов [2, 76]. Продолжая идею русских философов об избранности России, ее особой миссии, студенты писали о важности передачи культурных ценностей: *«Я считаю русскую культуру одну из величайших культур мира, которая несет всем народам духовность и гуманизм»* [2, 19].

На втором месте по частоте упоминаний находится наличие мифа об общих представлениях, национальный характер. Несмотря на то, что это понятие по-прежнему является одним из наиболее спорных, в понимании которого в научном обществе еще не пришли к единой точке зрения, почти половина студентов (49%) считают обязательным фактором принадлежности к своему этносу наличие особого образа мыслей, а также определенных черт характера: *«Здесь же можно отметить такое понятие, объединяющее меня с представителями*

¹ Архив этноавтобиографий Г.К. Уразалиевой.

русской нации, как менталитет. Менталитет я понимаю, это образ мыслей, мировоззрение, отношение к жизни и т.д.», — пишет студентка [2, 134]. Что касается черт характера, которыми обладает, по мнению студентов, русский человек, то наши результаты совпали с результатами З.В. Сикевич. Доброта, которую отметили 11% респондентов, оказалась на первом месте среди упомянутых качеств. Также упоминались такие черты русских людей как гостеприимство (9%), щедрость (7%), широта души (7%), отзывчивость (3%), трудолюбие (2%). *«Я русская, и это не национальность, это состояние души. И быть русским, значит быть человеком, который имеет широкую душу... Русские люди с их широтой души имеют такой набор индивидуальных качеств в своей крови, что никакой другой народ не может соревноваться с нами на этом поприще»*, — заключает девушка [2, 148]. Как и в исследованиях наших предшественников, «достоинства» встречались в ответах чаще, чем «недостатки», однако, студенты не отрицали наличия последних. К примеру, 9% студентов посчитали лень неотъемлемой чертой русского человека: *«Русские, по мнению большинства, считаются ленивой нацией, и с этим я частично соглашусь. На мой взгляд, лень — это один из основных недостатков, с которым необходимо бороться»*, — пишет студентка [2, 11]. Однако среди респондентов были и те, кто отрицал существование особого национального характера: *«Что такое вообще национальный характер? Есть определенный набор характеристик, которые мы привыкли давать тому или иному народу. Но это всего лишь те маркеры, которыми мы определяем свое ощущение самих себя или «других». Однако кто считал, сколько среди американцев расчетливых, среди французов обходительных, среди итальянцев вспыльчивых, среди русских хлебосольных? Эти штампы существенно сужают наш кругозор»*, — пишет студент [2, 1].

Еще одна важная особенность русских людей, на которую обратили внимание 11% респондентов, — это «загадочная русская душа». Студенты в своих работах писали о том же, о чем и Н.А. Бердяев: о специфичности русского характера, об удивительной русской душе, о ее антиномичности и противоречивости России в целом: *«Быть русским человеком значит иметь противоречивую душу, которая неподвластна логике, её можно лишь почувствовать»*, — рассуждает студентка [2, 38]. Другая девушка отмечает: *«Как это ни печально, быть русской, значит жить в стране с огромными просторами, богатейшими ресурсами и талантливейшими людьми и в то же время в стране бедной, разграбляемой и несчастной»* [2, 5]. Подтверждая слова русских философов и писателей, студенты писали, что только русские могут понять друг друга, понять противоречивость русской души и принять ее.

Следующий этнодифференцирующий признак, который был выделен нами как конструирующий признак (по В. Ильину) и подтвержден в ответах респондентами, — это идея общности происхождения и судьбы. 48% студентов причисляют себя к русским потому, что считают себя частью истории этого народа, принимают все прошлое, настоящее и будущее своей страны: *«На этой земле у нас общая судьба только со своим народом... История и культура нашего народа — это наша судьба и плоть нашего самоопределения»*, — пишет молодой человек [2, 130]. Кроме того, часть респондентов (7 %) считает, что помимо принятия истории своего этноса, необходимо нести ответственность за все события, которые произошли, а также за то, что происходит с русскими в настоящее время: *«Я — русский, и это значит, что за мной тысячелетняя история русской культуры, и это, кроме всего прочего, налагает на меня изрядную долю ответственности. Прежде всего, ответственности за собственное поведение, ответственность перед историей, в которой за право быть русским не раз приходилось отвоевывать у тех, кто жаждал полного истребления русского»*, — считает студент [2, 56].

Одна из причин причисления себя к русским, являлась этническая принадлежность родителей и прародителей. *«Я русская, для меня это значит, что мои родители русские, мои бабушки и дедушки тоже русские»*, — пишет одна из студенток [2, 111]. Другой респондент определяет свою принадлежность исключительно по родителям: *«Ведь национальность выбирается по родителям, поэтому моя национальная принадлежность обусловлена только этим»* [2, 19]. В большинстве работ подразумевалась идентификация по этнической принад-

лежности родителей, однако, напрямую это не говорилось. В основном это прослеживается в работах детей межэтнических браков, где студенты пишут, что являются, к примеру, наполовину русскими, наполовину армянами.

Более трети респондентов (37 %) объяснили свою принадлежность к русскому этносу тем, что родились в России, являющейся исторической родиной этого этноса: *«Я считаю себя русской, потому что, прежде всего я родилась на территории России»* [2, 37]. Одна из работ стала примером моноэтнической идентичности с чужой группой, когда произошло полное принятие норм и ценностей русского этноса: *«По национальности я к русской культуре не принадлежу, но, поскольку живу в России с рождения, то о своей этничности четкого представления не имею, поэтому отношу себя к русским»* [2, 82]. Почти половина опрошенных (47%) определили свою этническую идентичность через гражданство. *«Для меня быть русской — это прежде всего быть гражданкой Российской Федерации»*, — пишет студентка [2, 45]. Многие респонденты (14 %) акцентировали внимание на том, что родились и выросли именно в Москве, где уклад жизни и социализация личности имеют иной характер, нежели в других городах и регионах страны. Студентка пишет: *«Я больше считаю себя москвичкой, чем русской. Я понимаю, что это два абсолютно разных понятия, но русский человек, живущий в современной России в глухой тайге и русский человек, живущий в мегаполисе, не смогут, может быть, даже понять друг друга, потому что они разговаривают на разных диалектах, имеют разные жизненные ценности и цели, разную среду обитания. По сути, их объединяет историческое прошлое и общая территория, но этого недостаточно»* [2, 64].

42% студентов обязательным фактором причисления себя к русскому этносу назвали знание языка. Как уже ранее отмечалось, язык является одним из самых важных этнодифференцирующих признаков, он может, как усиливать ощущение своей принадлежности к этнической группе, так и заставлять сомневаться в своей идентичности, если язык не совпадает с языком этноса. Примером первого влияния языка являются следующие высказывания: *«Как я считаю, самым первым (не главным, а именно первым) критерием является русский язык. Именно в языке заключается народная мудрость, та, что существует испокон веков. Язык — это часть человека, это то общее, что есть у всего народа, что объединяет абсолютно разных и незнакомых людей в единое целое — народ»* [2, 22]. Однако есть примеры ситуаций, когда человек растет в межэтнической семье и знает несколько языков, в связи с чем появляется сомнение в этнической идентификации: *«Мои первые слова были на русском языке, но когда мне было 8 месяцев, мои родители решили расстаться... и я жила в мамином родительском доме, где говорили только на эстонском языке. Когда в 3 года меня отдали в обычный садик, я не умела говорить по-русски и не могла общаться с детьми, воспитателями. Тогда в экстренном порядке меня стали обучать русскому языку. Когда мне было 4 года, мои родители решили снова сойтись. Я много времени проводила с бабушкой по отцовской линии, а она говорила на украинском языке. Так получилось, что в 6–7 лет я свободно могла говорить по-эстонски и по-украински с моими родственниками... По моему внутреннему ощущению я не могу называть себя эстонкой, так же как я не могу назвать себя украинкой. А если по паспорту, то в нем у меня написано, что я русская»* [2, 133].

Еще один этнодифференцирующий признак — религия. Несмотря на то, что Россия — страна поликонфессиональная, часть респондентов (15%) отметили, что исповедование православия — характерная черта русского человека: *«Я русская, и это значит, что моя религия — православие. Для меня это незыблемый критерий, неоспоримая истина и назвать мусульманина русским я бы не смогла»* [2, 96].

Внешняя идентификация включена Ильиным в механизм социального конструирования этноса, однако, лишь 8 % опрошенных указали среди причин причисления себя к русским внешние данные (цвет волос, глаз и т.д.). *«Я — русская, это значит, что я обладаю характерной внешностью — у меня светлые волосы, светлая кожа и глаза»*, — пишет студентка [2, 84]. При этом нужно добавить, что внешняя идентификация особенно важна для детей межэтнических браков. Так, в результате травмы инаковостью (У. Эко) произошла этническая самоидентификация следующей студентки: *«Впервые моя национальность заинтересо-*

вала меня где-то лет в 9, когда другие дети стали спрашивать, кто я по национальности, говорили, что у меня «нерусская» внешность. То, что я наполовину татарка, на меня особо впечатления не произвело. Позднее, став старше, я начала стесняться того, что я не такая как все» [2, 137].

В 20 работах респондентов (11%) были отмечены стереотипные, упрощенные представления о характерных чертах своего этноса. В.И. Ильин отмечает, что именно стереотипы позволяют людям составить наглядную карту полиэтнического социального пространства, в котором они живут. Стереотипы формируют этническое поле как изнутри (автостереотипы), так и извне. В своих этноавтобиографиях студенты переносили на себя те стереотипные представления, которые характерны для их этнической группы. *«Я русская, а это значит, что мне хорошо, то немцу смерть»,* — замечает девушка [2, 112]. *«Русские в глазах иностранцев представляются близкими к медведям существа, в шапке-ушанке и бутылкой водки в руке. Такие клише, штампы не рождаются на пустом месте»,* — отмечает студент [2, 168].

Таким образом, студентами были затронуты, в первую очередь, элементы когнитивного компонента этнической идентичности, то есть они отмечали знания, представления об особенностях собственной группы и осознание себя как ее члена на основе определенных этнодифференцирующих признаков. Однако прослеживается и аффективный компонент идентификации, представляющий собой оценку качеств собственной группы, отношение к членству в ней, значимость этого членства. Результаты показали, что этнические чувства занимают второе место среди причин отнесения себя к русскому этносу. Так, 61% студентов говорили о гордости за свой этнос, за его историю и его достижения, декларируя, таким образом, позитивную моноэтническую идентичность: *«Я горжусь тем, что я русская. Я горжусь тем, что история земли русской так богата, как замечательными выдающимися людьми, так и тем, что эти люди прославили нашу страну, как одну из самых сильных держав на планете»,* — пишет девушка [2, 154]. *«... могу с уверенностью сказать, что я испытываю гордость за то, что являюсь частью такого народа. Гордость за тех евреев, что стали первооткрывателями и основоположниками многочисленных наук, тех, кто вносил немалую лепту в развитие искусства в нашей стране и в мире в целом»,* — отмечает студентка, ребенок межэтнического брака [2, 64]. Кроме того, в 24% работ говорилось о любви к своему этносу, к его культуре, достижениям и т.п. Студентка пишет: *«Я люблю Россию и русских людей, горжусь тем, что я родилась в России, рада, что мой муж русский и горжусь, что мои дети будут русскими»* [2, 114].

Однако помимо гордости за свой народ, некоторые студенты (8 %) испытывали стыд за свой этнос, разочарование в нем, демонстрируя свою негативную этническую идентичность: *«Быть русским значит быть несвободным, постоянно ставить себе все новые и новые рамки и границы. Жить в стране, в которой люди думают только о том, как бы больше заработать денег, потому что все (или очень многое) упирается в деньги... Быть русским — быть безучастным, верить телевизору, верить что уже ничего не изменишь, все воспринимать как данность, не подвергающуюся сомнению»,* — пишет девушка [2, 106]. Другой студент также говорит об упадке своего этноса: *«Быть русским значит, гордиться прошлым, а на настоящее закрывать глаза, когда заходит речь о русском духе; значит, что моя национальность растворяется в стране, где я родился, что расширяющийся гуманизм и тема моей работы не имеют точек соприкосновения в будущем»* [2, 113].

Как известно, несмотря на любые декларации, все этносы в полиэтнических обществах строятся на основе этноса-большинства в качестве ядра. В результате возникает некоторая иерархия этнического большинства-меньшинства: некоторые этнические группы (зачастую, титульные этносы) имеют какие-то привилегии в отличие от остальных. Такую же ситуацию отмечают респонденты нашего исследования. Студентка, идентифицировавшая себя русской, пишет: *«Быть русским в РФ выгодно с точки зрения политической обстановки. К сожалению (я не разделяю эту точку зрения), в РФ к русским относят очень лояльно, нежели к другим народностям, проживающим на территории РФ. Существует некоторое пренебрежение к культуре и традициям других народов, утробованный патриотизм перерастает в рус-*

ский национализм» [2, 95]. Другой студент отмечает: *«Быть русским в нашей стране – значит, не иметь практически никаких непреодолимых препятствий при трудоустройстве; ведь быть русским – значит, быть «своим», а «своим» у нас (так уж сложилось в стране с относительно недавних пор) – все со скидками»* [2, 26].

Итак, мы рассмотрели, какие когнитивные и аффективные компоненты включает этническая идентичность студентов. Как оказалось, в формировании этнического самосознания важную роль играет этническая осведомленность, знания о своей культуре, ценностях и традициях. Однако аффективный компонент имеет также большое значение, и более половины опрошенных студентов обладают позитивной этнической идентичностью. Далее мы перейдем к анализу условий формирования процесса этнической идентификации.

Пространственно-временной континуум формирования русской идентичности

Собранный массив данных позволяет прояснить, как в этноавтобиографиях задается пространственно-временной континуум самоидентификации. Под ним мы будем понимать неразрывность пространственных и временных характеристик протекающих реальных социальных процессов. Методика этноавтобиографий, помимо причин идентификации с определенным этносом, позволяет выявить, в какой период социализации и при каких условиях начинается формироваться этническая идентичность. Это является важным плюсом использования данной методики как дополнения к другим, чаще используемым методам.

Поскольку ранее было сказано, что далеко не все студенты затрагивали в своих этноавтобиографиях формирование этнической идентичности, лишь в четверти работ (23 %) упоминалось время осознания своей принадлежности к определенному этносу.

Начало формирования этнической идентичности, по данным этноавтобиографий, у некоторых респондентов (2 %) совпало с возрастом 3–4 лет. Однако связано это было, в первую очередь, с «травмой инаковостью» по этническому признаку. Одна из студенток вспомнила, что впервые столкнулась с человеком другой этнической принадлежности еще в детском саду, когда к ним в группу пришла девочка-мулатка, дочь русской женщины и африканского мужчины. И уже в том возрасте респондентка столкнулась с проявлением нетерпимости к этой девочке, что заставило ее задуматься, почему они отличаются друг от друга [2, 42].

Десятая часть респондентов (10 %) связали время своей идентификации со школьным возрастом, причем большинство из них (7 %) писали о возрасте 6–9 лет, то есть о начальной школе. Причины осознания своей этнической принадлежности в таком возрасте различны. Это могут быть как некоторые происшествия и изменения в структуре семьи и взаимоотношениях с ее членами, так и коммуникации с друзьями в школе и вне школы. Однако первостепенную роль на этом этапе играет семья и то, как именно она вводит ребенка в этнополе, как дифференцирует его этническую группу и чужую. Одна из студенток пишет: *«С детства я слышала слово “нацмен”, произносимое пренебрежительным, как бы «понимающим», но на самом деле неодобрительным тоном, каким говорят об алкоголиках из соседнего подъезда. И с самого детства слово “нацмен” ассоциировалось у меня с чем-то постыдным, плохим, чего не должно быть в нормальной жизни человека. Потом я поняла (лет в 7), что “нацмены” — это лица кавказской национальности. Когда я поняла, что “нацменами” мама называет таких же, как и мы, людей, то стала с ней спорить, доказывая, что они ничем не отличаются от нас»* [2, 94]. Данный пример демонстрирует трансформацию нормальной этнической идентичности в гиперидентичность со «своим» этносом в полиэтническом обществе. Такой тип идентичности сопровождается этноцентристскими стереотипами, предубеждениями по отношению к представителям других этнических групп (в данном случае, евреев и представителей кавказских этносов) и уклонению от тесного взаимодействия с ними.

Классик французской психологии Ж. Пиаже, приводя результаты своего исследования, говорил о том, что дети в начальных классах еще затрудняются соотносить части целого, одни предметы с другими, осознавать, что город может быть внутри страны. Дети этого возраста, как было отмечено ранее, еще плохо осознают свою принадлежность к этнической группе. Наше исследование также подтвердило это явление: *«Еще в начальной школе я писала сочинение о своей Родине, и тогда Родиной я назвала Москву. Однако потом учительница объ-*

яснила, что Родина — это больше чем просто город, это даже больше чем страна. А то, о чем писала я, оказалось маленькой Родиной: свой город, район, дворик... все то, что постоянно окружает меня, что является очень близким и дорогим» [2, 22]. Студент, ребенок межэтнического брака пишет: «Лет до шести я вообще не задумывался над тем, что есть какие-то различия между национальностями. Я видел пару раз негров, знал, что они далеко, в Африке, и для меня они как бы не существовали... Годом к восьми эта классификация расширилась: люди поделились на русских и евреев (видимо, потому что в моей семье о других этносах не говорили)... Только спустя несколько лет я узнал, что в России евреев лишь 0,5 % от общего числа населения» [2, 135].

Наше исследование показало, что для многих русских студентов этническая принадлежность является данностью, фактом. Это связано, в первую очередь, с воспитанием в русской среде, где твой этнос является доминирующим. 11 % опрошенных отметили, что впервые задумались о своей этнической принадлежности совсем недавно: «Я никогда не задумывалась над тем, русская ли я. Это не было для меня спорным вопросом. Для меня это естественно», — пишет девушка [2, 164]. Кроме того, оказалось, что задача написать этноавтобиографию в курсе «Этносоциологии» стала для 15 студентов нарративным толчком к саморефлексии своей этничности. Было отмечено, что только задание написать работу заставило их задуматься о своей этнической идентичности: «Получив задание написать этноавтобиографию, я столкнулась с серьезной проблемой, которая сводилась к вопросу «Что же писать?» Начав размышлять на тему «Я русская, а это значит...» я поняла, что всерьез не задавалась этим вопросом» [2, 75].

Как уже отмечалось выше, на детей и формирование их идентичности, в том числе этнической, самое большое влияние оказывают родители и семья в целом. Именно близкие вводят ребенка в этнополе, решают, где и как он будет расти и воспитываться, ценности какой культуры воспринимать и развивать. В работах трех респондентов развод родителей сказался на их осознании своей принадлежности к этносу. Так, одна студентка является ребенком межэтнического брака между армянином и русской: «Когда мне было восемь лет, мои родители развелись, и я осталась жить с мамой. На русской половине семьи. Наверное, можно сказать, что я больше русская, чем армянка. Я с восьми лет жила в Русской семье, с русскими мамой, бабушкой дедушкой, двоюродными братьями. Я не знаю армянских традиций и языка...» [2, 143]. Вторая студентка также пережила развод родителей, украинца и татарки, однако, в результате переезда в Россию стала считать себя русской: «Скорее всего, на мое представление о своей национальности влияет окружающая меня реальность. И, конечно же, важно в какой стране я живу, на каком языке я говорю и моя внешность. Вряд ли имей внешность узбечки, я бы считала себя русской. Сейчас я считаю себя русской и горжусь этим» [2, 144]. В результате мы видим, что сначала сформировалось маргинальная идентичность, когда девушка не усвоила ценностей ни одной этнической группы, а затем моноэтническая идентификация с этнодоминирующей группой.

В становлении этнической идентичности большую роль также играет школа: именно в ней ребенок познает свою культуру, учит язык, получает знания о своей стране. Несколько студентов отмечали, что введение некоторых предметов повлияли на осознание ими своей связи с русским этносом: «Когда я начала осознавать свою принадлежность к русской нации, я училась в школе. Мне так кажется, что сказалось влияние таких предметов, как география и история. География, прежде всего потому, что именно изучая этот предмет, по-настоящему понимаешь, каковы реальные масштабы твоей страны, какое место она занимает в мире. История — потому что осознаешь, какая богатейшая история у твоей страны» [2, 115]. Помимо учебы, в школе происходят и межэтнические коммуникации, когда дети различных этнических групп общаются друг с другом, таким образом, открывая для себя «другие» этносы, их культуру, ценности, еще больше расширяя свою этническую осведомленность: «До 14 лет я учился в довольно замкнутом внутри себя классе, где вопрос национальности не вставал. Для меня национальность является вещью необязательной, о которой можно и не знать и которой может не быть, а открытие факта дифференциации ко-

го-то из знакомых заставляло меня вновь присматриваться к этим людям – есть ли в них что-то, что отличает их от меня? Впрочем, так как эта тема меня никогда не заботила, я быстро бросал это занятие. Появление националистов в моей жизни совпало со сменой школы. Моя вторая школа была поистине интернациональной. Кто-то даже говорил, что у нас представлены 40 % кавказских народов. Были у нас и скинхеды» [2, 119].

Пространственный фактор также повлиял на становление этнической идентичности студентов. В процессе работы мы рассматривали фактор места рождения. Из 73 человек, указавших его в своих работах, 67 родились в России. Трое человек, которые местом рождения назвали страны СНГ, в настоящее время также живут в России. Исследование показало, что респонденты, которые переехали в Россию, сильнее ощутили связь со своим этносом: *«Мое ощущение того, что я русская, особо я стала осознавать еще с детства, т.к. родилась в Узбекистане и прожила там 16 лет. В 80-х-90-х годах в Узбекистане было много национальных «вспышек». Гонения турок, таджиков, в меньшей степени русских, естественно сказались на отношениях к другим нациям, живущим на территории Узбекистана. С самого детства, мое ощущение того, что я русская, носило характер униженной и презираемой нации. Очень часто высказывались в сторону русских слова о том, что Узбекистан для узбеков, а русские «убирайтесь» в свою Россию. Только переехав в Россию, я осознала, что имею такие же права, как и все живущие в России»* [2, 100].

7% респондентов отметили, что по-настоящему осознали свою этническую принадлежность, когда оказались за границей, среди других этнических групп. Одна студентка пишет: *«В 19 лет я вышла замуж за еврея, молодого человека с соседней улицы. Когда мы уехали в Израиль, все мои положительные качества вдруг трансформировались в мои недостатки. Находясь среди представителей другого этноса, я ощущала себя «белой вороной». Не было дня, чтобы мне в моей семье не напоминали, что я русская»* [2, 159]. Трое респондентов написали о том, что в период существования Советского Союза они ощущали себя советскими людьми, и так же их воспринимали и за границей: *«Когда мне исполнилось 9 лет, моих родителей послали в командировку за границу, и вот тогда я поняла, что люди бывают не только разных национальностей, но даже и разного цвета. Именно тогда я начала идентифицировать себя по национальному признаку — в отличие от местных жителей, а также других белых людей, я была СОВЕТСКОЙ. Не русской, а именно советской, т.к. в нашем посольстве работали люди совершенно разных национальностей. Наше посольство находилось рядом с американским и мы, дети, очень гордились, что живем именно в советском посольстве, и очень радовались, что наш красный флаг виден «более издалека», чем их американский»* [2, 75]. Как мы видим, осознание своей этнической идентичности складывается также в условиях межэтнических коммуникаций, причем конечный результат зависит от того, как именно воспринимают этническую группу человека в стране, где другой этнос является доминирующим. Однако исследование показывает, что, несмотря на пребывание за границей, самоощущение остается прежним. Несколько студентов (3%) считают, что никакая смена места жительства не влияет на этническое самосознание, подтверждая тем самым концепцию Ф. Барта о сохранении этнических границ: *«Границы государств могут меняться, приходят и уходят президенты, но никто не может повлиять на то, кем ощущает себя тот или иной человек. Место жительства не играет здесь существенной роли»* [2, 22].

Как мы выяснили, большинство респондентов впервые коснулись вопроса своей русской идентичности в детстве, причем некоторые еще в детском саду. Поскольку этническое самосознание непосредственно включено в процесс социализации ребенка, то его формирование напрямую связано с воспитанием в семье и школе. Большую роль в эти периоды играет иноэтническое окружение.

Значение иноэтнического окружения в процессе самоидентификации

При анализе роли иноэтнического окружения в процессе этнической идентификации студентов были выявлены следующие межэтнические коммуникации: внутри семьи, в школе, в компании друзей и за границей.

В первую очередь, мы рассмотрим условия формирования этнической идентичности респондентов, являющихся детьми межэтнических браков. В таких ситуациях ребенок растет на стыке культур двух и более этносов, и существует несколько путей развития этнического самосознания. Во-первых, это моноэтническая идентичность с одним из этносов. Как уже было отмечено ранее, 13 респондентов, являясь детьми межэтнических браков, выбирали русскую идентичность, объясняя это, в первую очередь, близостью русской культуры и родным русским языком. Во-вторых, это биэтническая идентичность, при которой человек в должной мере усваивает культуру обоих этносов. Из 24 респондентов, имеющих родителей разной этнической принадлежности, только одна студентка в своей работе показала, что обладает биэтнической идентичностью, поскольку в равной мере знает культуру двух этносов — русского и еврейского. Детство, проведенное в детских еврейских центрах, помогло ей узнать о еврейской культуре и традициях; жизнь в России дала ей знания о русском этносе, кроме того, она приняла православие [2, 119]. Однако и она, и еще 10 респондентов не смогли идентифицировать себя с какой-либо одной этнической группой. В нескольких работах студентами был использован термин «полукровка» для обозначения своей этнической принадлежности, или же использовалось сочетание названий двух этносов, например, «русский еврей» или «русская татарка». При этом если одни студенты все же осознавали свою идентичность с обоими этносами, то были и те, кто декларировал свою маргинальную идентичность: *«Я полукровка, и это значит — не чувствовать себя полностью представителем какой-то нации. Быть полукровкой — это иметь странное ощущение раздвоенности корней своего происхождения»*, — пишет студентка [2, 139].

Ранее уже отмечалось, что значимые проблемы семьи также сказываются на этносоциализации детей. Например, развод родителей приводит к тому, что ребенок из одной культуры переходит в другую, или же лишается окружения второй этнической группы.

Нами также анализировался концепт толерантности, без которой невозможно мирное сосуществование различных этносов, тем более в такой полиэтнической стране как Россия. Наши студенты также уделили внимание этой проблеме. Из 15 человек, отметивших равенство всех этносов, трое — дети межэтнических браков. Студентка, ребенок межэтнического брака отмечает: *«Быть полукровкой, на мой взгляд, значит не чувствовать себя ограниченными рамками каких-то определенных национальных традиций, быть более терпимым к другим нациям, чем «чистокровные» представители»* [2, 139]. Проблему нетерпимого отношения к людям другой этнической принадлежности в России отметила еще одна девушка, идентифицировавшая себя как «русскую татарку». Она мусульманка, однако, старается выглядеть европейкой, объясняя это следующим образом: *«Не стоит акцентировать внимание окружения на свою этническую принадлежность, если ты выросла и живешь в другой культуре. Я придерживаюсь правила «in Rome do as the Romans do», т.е. живи в Риме так, как это делают римляне. Если бы я жила на официальной мусульманской земле, где была бы принята основной мусульманская религия, возможно, я бы выглядела сейчас иначе»* [2, 105]. Однако русские студенты также обращали внимание на проблему толерантности, отмечая необходимость ее существования в современном обществе: *«Только познав себя, можно дойти до того уровня, чтобы иметь способность познавать и понимать других. В этом смысле толерантность — огромный труд. Это не витающий где-то в научных кругах термин, украденный у биологов. Толерантность — это не терпимость ни в коем случае и не терпение, а трудоемкий процесс самопознания на первом этапе и самоопределения через знание других впоследствии. Поэтому я считаю глупо кричать на каждом углу «Россия для русских!», когда не понятно, а для каких русских»* [2, 81].

Наличие иноэтнического окружения непосредственно сказывается на формировании этнической идентичности ребенка. Если человек растет в полиэтническом обществе и в процессе социализации имеет опыт межэтнического общения, то у него складываются нормальные взаимоотношения с другими этносами, при этом он обладает позитивной этнической идентичностью. Мы также выяснили, что в ситуациях, когда ребенок рождается в межэтническом браке, он оказывается посреди нескольких этнических культур, что затрудняет его

этническую самоидентификацию, однако, размытость этнокультурных границ делает его более толерантным по отношению к другим этносам.

Роль значимых взрослых в формировании русской идентичности

Методика этноавтобиографий позволила нам выявить, какое место в формировании этнической идентичности занимали значимые в жизни детей взрослые, такие как родители, воспитатели, учителя. Как писали П. Бергер и Т. Лукман, эти значимые другие играют роль посредников между детьми и объективным социальным миром. Именно то, как определяют ситуацию близкие люди, какие значения вкладывают в происходящее вокруг, является для ребенка объективной реальностью. Взрослые в данном случае «выбирают те или иные аспекты этого мира в зависимости от того места, которое они занимают в социальной структуре, и от своих индивидуальных, биографических характерных особенностей»¹ Таким образом, социальный мир предстает перед ребенком в модифицированном значимыми другими виде.

Как уже было сказано ранее, 47 % опрошенных студентов отметили этническую принадлежность родителей как один из основных факторов своей самоидентификации. В семье закладываются основные правила поведения в своей и в других этнических группах, а также культурные ценности, ритуалы, обычаи и традиции. Это, в первую очередь, происходит по принципу подражания, когда родители воспринимаются как образцы поведения данной культуры. Кроме того, взрослые вводят ребенка в этническое поле, прививая любовь к этнической литературе, музыке, традиционным праздникам и т.д. *«Родители с детства старались показать мне русского человека, прежде всего с помощью детских книг или каких-либо рассказов, в которых отлично описан этот образ»,* — пишет студентка [2, 147].

Крайне важно, чтобы взрослые стимулировали интерес ребенка к собственной этнической группе, к ее истории, конкретно истории их семьи. Именно межпоколенная передача ценностей имеет особое значение для воспроизводства этничности и формирования этнической идентичности. Так одна из студенток отмечает, что именно от деда она получила знания о своих корнях: *«Он [дед] рассказывал мне о том, что наша фамилия, Науменко, — это очень старинный украинский род и одна из самых распространенных фамилий, что портрет какого-то его деда или прадеда висит в местном музее, потому что он был одним из долгожителей того времени»* [2, 96].

Поскольку именно взрослые люди, родители являются основными агентами социализации в детстве, любое их мнение, утверждение, поведение сразу же отражается в сознании ребенка. Так, если родители настроены против некоторого явления, то и ребенок растет с негативным отношением к этому предмету. Примером является девушка, в своей этноавтобиографии рассказавшая, как родители с самого детства прививали ей нелюбовь к евреям и представителям кавказских этносов: *«Мама, рассказывая мне, кем были мои предки, чем они занимались, несколько раз повторяла, что кого-кого, а евреев в нашем роду нет; якобы, это специально блюлось. И просила меня ни в коем случае не выходить замуж за еврея, а тем более за кавказца. Говоря о кавказцах, она употребляла слово «черный». Когда я спрашивала: «Мама, а как же греки, итальянцы?», она отвечала, что они те же самые «чурки». Признаться, это слово резало мне слух... Сейчас мне почти 21 год. 21 год моей русской истории и около 15 лет споров до хрипоты с близкими людьми. И что самое смешное, я четко осознаю, что не хочу связывать свою жизнь с кавказцем или евреем»* [2, 94]. В случае если родители не делают акцента на этнической принадлежности ребенка, то он о ней не задумывается, а принимает как должное: *«Родители никогда не говорили со мной о том, к какой национальности мне следует относить себя. Я не испытывала в этом плане давления социального окружения. Естественным образом все сложилось так, что мне стало органично считать себя русским человеком»* [2, 24].

Помимо семьи, на формирование у детей этнического самосознания могут оказывать влияние и другие люди. Так, например, одна из студенток вспоминала: *«Дело в том, что в*

¹ Бергер П., Лукман Т. Социальное конструирование реальности: Трактат по социологии знания М.: Академия-Центр; Медиум, 1995.

возрасте 7 лет погиб отец и моей маме пришлось выполнять обе функции родителя. Вследствие того, что мама была в постоянных разъездах по работе, моим воспитанием занималась лучшая ее подруга, русская по национальности. Несомненно, это очень сильно повлияло на мой быт, который со временем утратил мусульманскую окраску совсем» [2, 105].

Студенты также отмечали, что знания о русском этносе они получили в школе, а учителя привили им толерантное отношение к людям. *«Большую роль в воспитании во мне русского человека сыграла школа, учителя, изучение русской истории, потому что именно в школе отвечают на многие вопросы, непонятные ребенку»*, — пишет студентка [2, 114]. Другая девушка, ребенок межэтнического брака, замечает: *«Наверное, благодаря классному руководителю, учителям и школе в целом, вопрос о национальной принадлежности учеников не возникал»* [2, 105].

Таким образом, исследование показало, что именно взрослые в большей степени влияют на формирование этнических чувств детей, именно на них ориентируются дети при выборе своей этнической идентичности. Кроме того, от воспитания зависит и то, насколько толерантен будет ребенок, как именно будут происходить его межэтнические коммуникации.

Оценка этноавтобиографической методики студентами

Ранее мы выяснили, что для 15 студентов задание написать свою этноавтобиографию стало нарративным толчком к саморефлексии. Кроме того, некоторые респонденты отмечали важность написания такой работы, поскольку это позволило всерьез задуматься о своей этнической принадлежности, причем многим, как оказалось, впервые. В связи с этим, в апреле-мае 2009 года одним из авторов, Е.В. Беккеревой, было взято 20 полуформализованных интервью у студентов 4–5 курсов социологического факультета РГГУ, которые годом-двумя ранее писали свои этноавтобиографии. Наша задача была выяснить, как сами студенты оценивают такую методику в курсе «Этносоциологии», позволила ли она им узнать что-то новое о себе и своей идентичности.

Первый вопрос, который был задан респондентам, — помнят ли они тексты своих работ или хотя бы основные мысли из них. Мнения разделись почти поровну: 9 респондентов в точности вспомнили то, о чем писали, а 8 респондентов не смогли сформулировать даже направление своих рассуждений. Еще трое человек высказали, по их мнению, те идеи, речь о которых шла в их работах, однако, после проверки текстов оказалось, что это не совсем так. Один из студентов, родившийся в Крыму и проживший там до окончания школы, думал, что также написал и про переезд в Москву, который усилил его русскую идентичность, однако про этот факт в работе ничего не было сказано [2, 165]. Это позволяет предположить, что осознание важности смены места жительства для этнической самоидентификации пришло несколько позже после написания задания. Другая студентка была уверена, что писала о заслугах России, о спортсменах, культурных деятелях и проч., благодаря которым хочется гордиться тем, что ты русский [4, И-3]. Однако в ее этноавтобиографии такой мысли не было [2, 16], а значит, теперь она выделяет еще одну причину причисления себя к русскому этносу. Надо сказать, что на вопрос о том, что может она сказать о своей этнической принадлежности теперь, спустя два года, студентка изложила основные идеи своей работы. Что касается других респондентов, только пять человек рассказали о каких-то изменениях в своей русской идентичности. Двое из них отметили, что за это время ощутили большую связь с русским этносом. Одна девушка рассказывает: *«По мере расширения границ опыта уровень рефлексии над своей национальной идентичностью в моем случае повышается. Я больше внимания стала уделять ее последствиям для моей реальной жизни»*¹ [4, И-4]. Другой респондент отмечает: *«Мое понимание того, что я русский, лишь углубилось по мере того, как я начал осознавать, насколько точно я воспроизвожу некоторые противные мне черты русской культуры»* [4, И-5]. Остальные опрошенные студенты ничего нового о своей идентичности сказать не смогли, подтвердив только ее связь с русским этносом.

¹ Архив Беккеревой Е.В. Интервью об этноавтобиографиях

Теперь перейдем к вопросам, которые непосредственно касаются методики этноавтобиографий. Большинство студентов (15) методику этноавтобиографии посчитали богатым материалом для изучения, поскольку он дает новое знание об особенностях этнической идентичности. Что касается пользы для самих респондентов, то, по их мнению, написание этноавтобиографии побуждает к саморефлексии, анализу своей этнической принадлежности. Одна из студенток говорит: «...[Методика] заставляет респондента не просто машинально отвечать на предложенные вопросы, но рефлексировать, размышлять комплексно, дает возможность взглянуть на процесс собственного становления через призму национального самосознания» [4, И-2]. Другая девушка отмечает: «Мне кажется, она эффективна, поскольку человек напишет о том, что ему действительно важно, а не то, о чем его спросит исследователь» [4, И-4]. Еще одно достоинство методики — ее свободная форма изложения, не имеющая каких-либо ограничений. Студенты считают, что такого рода творческие задания должны быть включены в программы и других дисциплин. Надо сказать, что в самих этноавтобиографиях респонденты также отмечали свой интерес к полученному заданию: «Честно говоря, я был очень рад получить такого рода задание, потому что в повседневной жизни редко предоставляет повод рассмотреть себя с точки зрения принадлежности к какому-либо этносу» [2, 135]. Среди плюсов написания работы студенты также выделили возможность оформить в текст те мысли, которые у них возникали ранее по поводу своей этничности. Однако многие респонденты столкнулись с некоторыми трудностями при работе. В одной из этноавтобиографий студентка писала: «Несмотря на то, что задание очень интересно, его очень сложно выполнить: я считаю, что вложить в текст все то, что я испытываю в отношении своей национальности, практически невозможно» [2, 47]. Главная же проблема, которая возникла перед студентами, — определение тех признаков, по которым они себя идентифицируют как русских: «Трудность была в том, о чем я косвенно уже сказала: надо осмыслить то, что ты воспринимаешь как само собой разумеющееся» [4, И-4].

Итак, мы выяснили, что, несмотря на вызывающий интерес, методика этноавтобиографий ставит перед студентами нелегкую задачу: разобраться в особенностях своей идентичности.

Webanalyst как инструмент для автоматического смыслового анализа текста

Харламов Александр Александрович,
ООО «Научно-производственный инновационный
центр Микросистемы»,
Сергиевский Николай Александрович,
Федеральный институт развития образования

Обработка больших массивов неструктурированной текстовой информации в настоящий момент является одним из наиболее востребованных направлений в области интеллектуальных технологий. Однако, не так много продуктов, которые оказываются реально полезными в прикладных исследованиях, связанных с обработкой текстовой информации. Одним из таких продуктов является последняя разработка в линейке продуктов ряда TextAnalyst. WebAnalyst представляет собой платформу для обработки текстовых массивов с использованием web интерфейса. Целью разработки является предоставление среды с эффективным набором инструментов отображения, поиска и обработки текстовой информации. В отличие от более ранних продуктов ряда, описываемая система позволяет хранить и обрабатывать текстовую информацию в наиболее распространенных форматах (txt, html, pdf, doc). Также система предусматривает построение нескольких типов отображения семантической сети. Единая среда хранения данных позволяет накапливать статистику и производить статистические выводы за счет пользовательских манипуляций. Система реализована на основе ядра TextAnalyst SDK. Основные возможности, предоставляемые пользователю программной системой WebAnalyst:

- анализ содержания текста с автоматическим формированием семантической сети с гиперссылками: получение смыслового портрета текста в терминах основных понятий и их смысловых связей;

- анализ содержания текста с автоматическим формированием тематического дерева с гиперссылками: выявление семантической структуры текста в виде иерархии тем и подтем;
- поиск релевантного ответа на запрос пользователя в неструктурированных текстах;
- автоматическое реферирование текста: формирование смыслового портрета текста в терминах наиболее информативных фраз;
- кластеризация информации: анализ распределения материала текстов по тематическим классам;
- автоматическая индексация текста с преобразованием в гипертекст;
- ранжирование всех видов информации о семантике текста по «степени значимости» с возможностью варьировать детальность ее исследования;
- автоматическое/автоматизированное формирование полнотекстовой базы знаний с гипертекстовой структурой и возможностями ассоциативного доступа к информации;
- работа пользователя в интерактивном режиме с целью настройки системы на конкретную предметную область, дополнения и корректировки результатов извлечения информации;
- web-интерфейс позволяет создавать личное пространство пользователя, доступ к которому пользователь может получить с любой точки доступа;
- предоставление инструментов социального взаимодействия пользователей;
- хранение структурированной информации пользователя в том числе: текстов, медиаданных и личных заметок.

Функциональность WebAnalyst'a

Формирование ассоциативной сети. В основе обработки и представления информации в системе WebAnalyst лежит функция автоматического восстановления внутренней структуры анализируемого текста. Из текста автоматически извлекаются числовые характеристики ключевых понятий и связей между ними. На основе этой информации строится ассоциативная сеть наподобие тех, что рисуют психолингвисты иллюстрируя структуру семантических представлений у человека. Визуализированная ассоциативная сеть позволяет познакомиться с основным содержанием предметной области, охватываемой текстом. Сочетания понятий, соединенных связью, вызывают у пользователя совершенно конкретные ассоциации, так как именно сочетания слов содержат основную информацию о модели мира. Таким образом, ассоциативная сеть, или структурный портрет текста, вводит пользователя в предметную область, описываемую текстом.

Формирование тематического дерева. Если из ассоциативной сети извлечь минимальный древовидный подграф и упорядочить его вершины по весам, получится так называемое тематическое дерево. Это другой способ познакомиться с содержанием текста, более привычный пользователю. Итог напоминает оглавление, в котором заголовки первого уровня относятся к главам, второго уровня — к параграфам и т.д. Тематическое дерево тоже позволяет пользователю представить себе предметную область, которая описывается текстом.

Формирование гипертекстовой структуры и средства навигации по ней. Поскольку мы имеем исходный текст и некоторое средство, сжато описывающее этот текст, мы можем, установив между одинаковыми понятиями гиперссылки, автоматически сформировать гипертекстовую структуру. На этот инструмент мы можем навешивать дополнительные функции, например, можем ставить в соответствие каждому понятию сети множество предложений, их содержащих. Или каждой паре понятий. Это уже существенно более мощный инструмент. Используя гипертекстовую структуру, пользователь может знакомиться с текстом, обращаясь к контексту всякий раз, когда требуется более детальное представление о мысли автора. При этом ассоциативная сеть или тематическое дерево служат средством навигации по тексту. Первоначальное знакомство с понятиями текста осуществляется по ассоциативной сети или тематическому дереву, дальнейшее знакомство с текстом — отсылкой к предложениям текста, в том числе в самом тексте, содержащем эти понятия.

Автоматическое реферирование текста. Эта функция могла бы числиться первой, поскольку позволяет познакомиться с текстом быстрее всего, однако ее можно реализовать только после создания ассоциативной сети. Ассоциативная сеть, т.е. ключевые понятия и их связи с их числовыми характеристиками, фактически формируют метрическое пространство, в котором и весь текст, и каждое его отдельное предложение являются вектором. Проранжировав предложения текста по весу (по длине вектора) и убрав предложения с весом меньшим порогового значения, мы можем получить что-то вроде реферата, если расставим оставшиеся предложения в порядке их появления в тексте.

Наряду с обычным рефератом можно построить так называемый тематический реферат. Для этого из ассоциативной сети выделяется подсеть, описывающая заданную запросом тему. Все предложения текста, содержащие понятия этой сети и имеющие вес выше порогового, составляют тематический реферат.

Сравнение и автоматическая классификация текстов. Данная функция реализована технологией TextAnalyst SDK, но не включена в функционал WebAnalyst'a. Можно сравнивать близость двух текстов, вычисляя степень корреляции между их ассоциативными сетями. Для этого вычисляется вектор пересечения этих текстов в декартовом произведении пространств каждого из текстов. Если имеется несколько тематических классов, входной текст может быть автоматически отнесен к одному из этих классов вычислением степени корреляции сети входного текста и сетей каждого тематического класса.

Автоматическая кластеризация текстов. Данная функция также реализована технологией TextAnalyst SDK, но не включена в функционал WebAnalyst'a. Если разбить сеть на подсети удалением слабых связей, можно предположить, что подсети описывают некоторые предметные области — подобласти предметной области текста. Понятие, имеющее наибольший вес в данной подсети, обозначает данную предметную подобласть. Если мы имеем несколько документов, разбивая их общую сеть на части, мы можем отнести эти документы к соответствующим подсетям, сравнивая сети документов с сетями предметных областей.

Формирование персональной базы знаний. Используя вышеперечисленные функции можно обрабатывать сколько угодно текстов и порознь и вместе. Результаты такого анализа можно сохранить, формируя собственную базу знаний. Можно иметь столько папок, сколько предметных областей, их разделов, их объединений предпочитает иметь пользователь.

Редактирование ассоциативной сети и тематического дерева. Ассоциативная сеть, также как и тематическое дерево, может быть отредактирована. В ней можно заменить название понятия, исключить его и исключить связь.

Используя перечисленную функциональность, автор или редактор е-книги легко, в течение двух недель, может подготовить свою базу, а затем, зашить ее в оболочку. Для хорошего качества анализа необходимо подбирать материал таким образом, чтобы он был тематически однородным: в случае книги материал надо разбивать на главы; также можно объединять несколько тематически однородных статей.

При желании можно навесить на соответствующие объекты текста (слова, предложения) информацию других медийных модальностей: аудио, видео. Точно также можно иметь параллельно гипертекстовому представлению плейн текст со всем его оформлением, с прямой ссылкой из гипертекста в текст и обратно.

Способы отображения ассоциативных сетей, полученных с помощью WebAnalyst

Необходимо расположить понятия ассоциативной сети на плоскости таким образом, чтобы их расположение коррелировало с их связями между собой. При решении этой задачи необходимо помнить, что человек не в состоянии следить за слишком большим количеством объектов. Поэтому при отображении большого числа понятий, может наблюдаться «замусоривание» поля малозначащими понятиями, что будет осложнять навигацию.

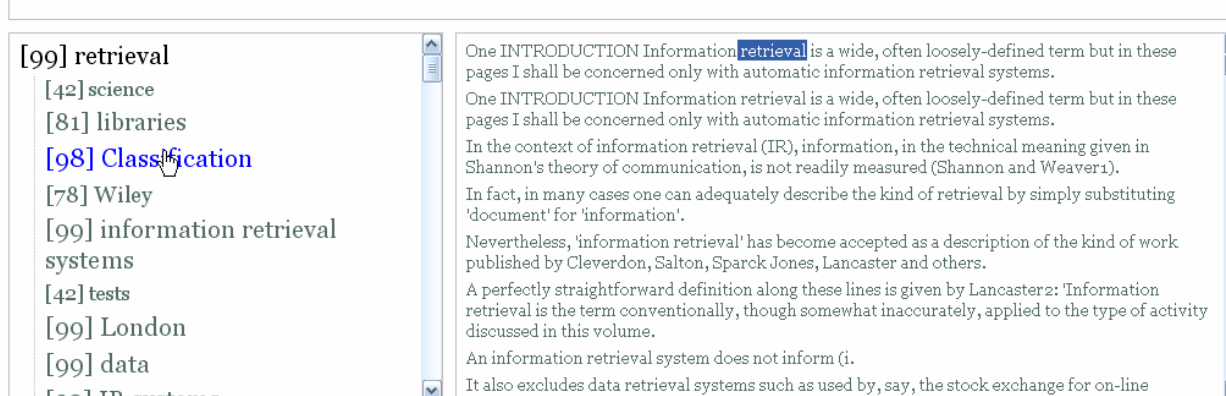


Рис 1. Интерфейс просмотра аналитического представления текста
(на примере курса английского языка).

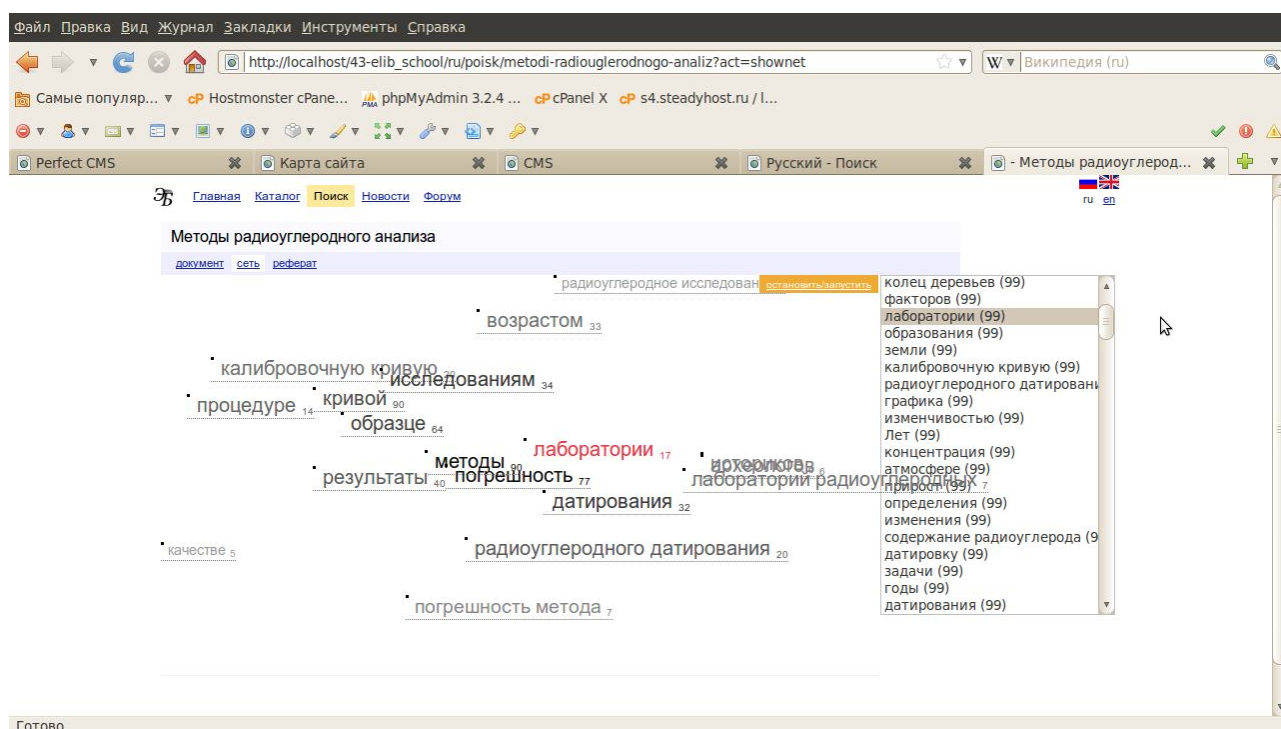


Рис 2. Интерфейс просмотра семантической сети в виде «звездного неба».

Дерево. Данный алгоритм изначально реализован в приложении TextAnalyst. Визуально он представляет собой обычные деревья, каждая ветка которых представляет собой список понятий, связанных с корневым понятием для этой ветки. Понятия сортируются по степени убывания их веса. В список первого уровня попадают понятия, имеющие вес не менее некоторого, заранее заданного, порога. Несомненные плюсы этого алгоритма в его простой реализации и привычности вида; кроме того, за счет сортировки и отсеивания по весу для первого уровня в поле зрения оказывается удобное для восприятия количество понятий.

«Звездное небо». Можно сформулировать следующее правило расположения понятий: «Расстояние между любой парой понятий должно быть обратно пропорционально весу связи между ними». Данный вид отображения сети может более наглядно отобразить многомерное пространство связей и понятий на 2-х или 3-х мерную плоскость, что позволяет пользователю лучше понять структуру анализируемого текста.

Дополнительные возможности

Система построения и анализа тестов. Очевидно, что электронные тесты должны подразумевать автоматическую проверку и выдачу пользователю только оценки в виде «да-нет».

Это справедливо для простого тестирования, построенного на выборе из нескольких вариантов одного правильного. Однако можно предложить и более сложные модели тестирования, автоматическая проверка которых является отдельной задачей.

Введено слов: всего - 0, верных - 0, неверных - 0
Открыто 0% текста

Проверить слово

Название текста

Share pictures ***** , ***** a ***** pictures *****
***** , ***** , *****
***** , *****
***** , *****
***** the *****

Рис 3. Пример теста «восстановление текста».

В рамках данной системы реализовано несколько типов тестов, так называемых текстовых манипуляторов. Все они ориентированы на обучение иностранному языку и служат как средством проверки, так и средством обучения. Предлагаемый продукт, таким образом, является не только средством для формирования, анализа и использования базы текстов, которая, фактически (за счет применения средства навигации по тексту его смыслового портрета — ассоциативной сети) превращается в базу знаний, но также является удобным электронным цитатником.

Прогнозирование социально-экономических процессов на основе анализа СМИ и социальных медиа

Чесноков Игорь Алексеевич,
ООО «Интегрум Медиа»

Ключевой тренд современного развития — информатизация, подразумевающая последовательное повышение насыщенности информационного фона, рост скорости распространения информации, а также её общую ценность как ресурса для обеспечения базы для принятия решений.

Социальная и экономическая активность гражданина современного государства постепенно смещается в информационное поле, что обусловлено развитием цифровых технологий и медиарынка в целом. Информатизация коснулась всех без исключения сфер жизнедеятельности общества: работа и бизнес, обучение, личное общение, информирование и потребление официальной информации. Не остались в стороне и нелегальные виды активности, преступность. Информатизация и виртуализация процесса возникновения и развития разнообразных социально-экономических явлений привела к фактическому замыканию таких активностей на современную информационную среду. В настоящее время возникновение нового «Майдана», «Манежки» и т.п. без использования традиционных медиа, а также социальных сетей, форумов и блогерских сервисов практически невозможно.

Эксперты по анализу информации и инициаторы разнообразных акций склонны разделять и противопоставлять друг другу социальные (блоги, социальные сети, совмещенные системы) и «традиционные» (телевидение, печать, радио, электронные СМИ) медиа. Это можно считать заблуждением с профессиональной точки зрения, поскольку информационное пространство следует изучать как единую среду, если речь идёт о необходимости своевременно и эффективно исследовать главенствующие тренды. Данное утверждение справедливо и в случае, когда возникает потребность в прогнозировании и предупреждении социально-экономических явлений.

Ожидается, что все большее значение будут приобретать системы перекрёстного анализа данных, позволяющие синхронно изучать тренды, отражённые в обеих составляющих национальной информационной среды. Как только будет доступен эффективный анализ информаци-

онных трендов местного (странового) масштаба, вероятно, произойдет активная интеграция с модулями изучения международных тенденций. В качестве примера можно привести недавние события в Тунисе, где падение режима привело к дестабилизации обстановки на территории сопредельных и находящихся в единой культурно-этнической группе государств.

На этом фоне повышается сложность стоящих перед аналитическими и экспертными службами задач. Сложность преимущественно возрастает за счёт увеличения количества изучаемых источников, что повышает неструктурированность изучаемой информации. В настоящее время эффективное изучение социально-экономических трендов, помимо экономических показателей на уровне региона, систематический сбор публикаций и телепрограмм дополняется учетом активности пользователей социальных сетей. Решение поставленных задач требует комплексного подхода, который, в свою очередь, подразумевает под собой использование аналитического инструментария в зависимости от типа информации, которая анализируется и готовится к сопоставлению.

К числу наиболее распространенных инструментов, использующихся при изучении данных СМИ целесообразно отнести мониторинг, контент-анализ, оценку тональности, а также ретроспективный анализ. Такое сочетание аналитических инструментов позволяет получить максимально объективную информацию о социально-экономических трендах, протекающих в обществе. Дополнительным условием получения объективных выводов является проведение комплекса мероприятий, направленных на выявление ангажированности конкретного источника контента. Такие мероприятия могут включать в себя кабинетную работу, подразумевающую установление аффилированности СМИ, а также полевую работу с непосредственными авторами того или иного материала.

Отметим, что текущий уровень развития технологий позволяет в значительной степени автоматизировать процесс массовой обработки информации, поступающей из «традиционных» СМИ. В целом, качественный комплексный анализ главенствующих в национальной медиасреде тенденций требует использования человеко-машинных систем. В особой степени данная потребность проявляется при анализе информации из социальных медиа, в частности, для определения качественных характеристик информационного фона.

Для этого мониторинг и контент-анализ дополняется психолого-лингвистической обработкой текстов. В анализ включаются графические, фото и видео материалы, публикуемые в социальных медиа в рамках определенного тренда. В результате исследователь получает возможность оперативного делать сопоставления по следующим параметрам:

- упоминаемость темы (персоналии) в СМИ в динамике;
- характер (тональность) публикаций;
- состав и структура (в т.ч. территориальная) задействованных СМИ;
- финансовые отношения источника материалов;
- политическая ангажированность источника/автора материалов;
- упоминаемость темы в социальных медиа;
- социальная принадлежность задействованных граждан;
- идентификация задействованных организаций;
- стихийное/организованное формирование тренда;
- динамика социально-экономических показателей региона.

В зависимости от специфики изучаемого тренда набор параметров может изменяться.

Таким образом, проблема анализа неструктурированной информации в социологических и смежных исследованиях решается посредством унификации исследовательских методов и совмещенным применением программно-аппаратной и экспертной обработки поступающих данных. Комплексный подход, основанный на синхронном применении представленных выше возможностей, на текущем уровне развития технологий альтернатив не имеет.

Архитектура системы динамического анализа неструктурированных данных

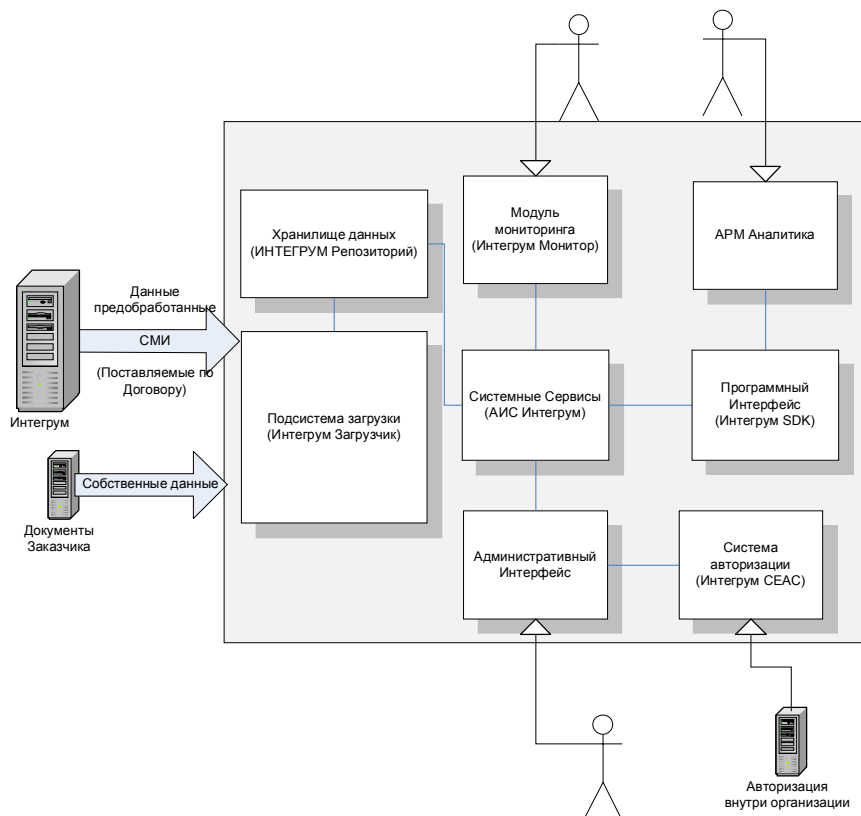
В настоящее время на рынке представлены следующие классы аналитических инструментов:

- системы сбора и архивирования документов Интернет и СМИ (агенты-роботы сборки

данных (краулеры), документные хранилища);

- конвертеры формата документов (Microsoft Office, Adobe Acrobat PDF, и т.п.);
- модули лингвистического анализа;
- системы контекстного поиска;
- системы анализа массива данных, использующие различные алгоритмы и модели математической обработки, позволяющие решать аналитические задачи, выявлять и отслеживать информационные объекты террористической направленности;
- системы выявления в текстовой информации объектов различного типа (персона, организация, географическое понятие, событие);
- системы семантического анализа (выделение роли и связей объектов), с целью выявления контента террористической направленности и связанных с этим объектов исследований, содержащих заданные признаки критических событий;
- системы оценки и прогноза развития социально-динамических процессов;
- инструменты настройки ситуационных моделей, применяемых для осуществления прогноза развития событий;
- экспертные системы поддержки принятия управленческих решений.

Комплексное применение вышеуказанных разработок позволяет автоматизировать оперативное выявление кризисных явлений и ситуаций и выработать обоснованные рекомендации для принятия управленческих решений, предложить готовые поведенческие модели при выявлении объективных признаков наступления (прогноза) социальных, этнических, конфессиональных, политических, экономических конфликтных ситуаций.



К настоящему моменту уже сложились особые требования к производительности математических моделей и СПО, а также к структуре хранилища для обеспечения оперативной и надежной обработки, непрерывно поступающих информационных потоков, соответствующих нашим внутренним требованиям, связанным с ежедневной обработкой и анализом огромных массивов данных. Все рабочие места для специалистов Заказчика обычно объединяются в единую инфраструктуру, позволяющую организовывать как независимую, так и «командную работу» аналитиков при сохранении максимальной гибкости конечных настроек для каждого рабочего места.

Подсистема	Назначение
ИНТЕГРУМ ЗАГРУЗЧИК	- Получение данных из разнородных информационных источников в едином формате; - Индексирование и загрузка данных; - Контроль получения и обработки данных.
ИНТЕГРУМ РЕПОЗИТОРИЙ	- Ведение реестра обработанных документов - Выдача документов по параметрам выпуска и источника
ИНТЕГРУМ ПРОФИ	- Предоставление пользовательского интерфейса для поиска, выдачи списка результатов и найденных документов; - Дополнительные формы для специализированного поиска; - Управление отчетами.
ИНТЕГРУМ МОНИТОР	- Формирование пользовательского интерфейса для мониторинга объектов по атрибутам; - Создание объектов для мониторинга (Название, ключевые слова, слова исключения, запрос); - Выполнение запросов для формирования статистики (Динамика сообщений); - Представление результатов мониторинга в виде гистограмм, графиков (по датам, по регионам, по категориям СМИ); - Сохранение результатов анализа.
АРМ АНАЛИТИКА	Осуществление мониторинга и контент-анализ сообщений СМИ с целью анализа и прогнозирования социально-политических, социально-экономических модернизационных трендов.
ИНТЕГРУМ АДМИНИСТРАТОР	- Управление набором категорий, источников и правами доступа к ним; - Управление правами доступа пользователей к наборам источников.
ИНТЕГРУМ SDK	Предоставление машинного доступа для операций поиска, вывода списка результатов и получения документов.
ИНТЕГРУМ SEAS (Система единой авторизации и синхронизации)	Обеспечение автоматической авторизации пользователей интегрируемых приложений для исключения необходимости повторной авторизации.

Комплексное, сбалансированное решение стоящих перед Заказчиком задач достигается за счет использования ранее созданных специализированных модулей и подсистем, которые функционально расширяются и настраиваются под задачу. Состав и назначение компонентов автоматизированной информационной системы (АИС) отражены на схеме и в таблице.

Любые характеристики контента могут быть приписаны как каналу информации, так и объекту освещения, поскольку отражают взаимоотношения между ними; т. е. они в равной степени характеризуют и поведение канала, и положение объекта. Поведение канала описывается с помощью таких категорий, как «выбор», «освещение», «внимание», «упоминание», «характер упоминания», «предпочтение», «акценты», «способы отражения», «способы проявления внимания» и т.п.

Исходные принципы стандартизации контент-анализа СМИ

Положение объекта описывается с помощью категорий «присутствие», «упоминаемость», «включенность», «доминирование», «представленность» и т.п.

Клиентов, представляющих маркетинг, PR, политтехнологи, как правило, интересует положение объекта; категориальный аппарат анализа формируется исходя из интереса клиента.

Часть единичных признаков поведения канала или положения объекта не поддается количественному измерению, поскольку речь идет о формах и способах освещения (проявления внимания, присутствия и т.п.).

К числу широко используемых качественных признаков относятся способ упоминания/освещения объекта (простое упоминание, главная тема и т.п.); знак отношения коммуникатора и других субъектов выраженного мнения к объекту; окраска контекста, в котором обнаруживается упоминание объекта (нейтральная, конфликтная, чрезвычайная и т.п.). Качественная шкала способов присутствия объекта в контенте должна содержать, как минимум, следующие пункты (по нарастанию присутствия):

- простое упоминание/называние — отдельно или в перечислении подобных объектов;
- высказывание (в том числе цитирование) в пределах одного абзаца основного текста или одной сюжетной единицы радио-, телесюжета¹;
- описание/оценка объекта как одна из тем публикации, сюжета, передачи; фотоиллюстрация с присутствием объекта на первом плане среди других;
- вынесение имени/названия/события в заголовок, лид, шапку полосы, анонс/ «шпигель» новостного выпуска;
- главная тема материала (не менее половины объема текста или времени).

Степень/меру присутствия можно определять в пространственно-временных единицах или приписывания весовых баллов экспертами.

При разработке сложных индексов, отражающих поведение канала или положение объекта, возможен отдельный учет особо акцентированных форм: внимание к объекту, проявленное в целиком посвященных ему материалах, полосах, разворотах, выпусках, или вынесение в заголовок, лид, «шапку» полосы. Например, можно использовать индекс, отражающий меру и характер внимания (индекс меры внимания): доля материалов, где объект присутствует в качестве главной темы по отношению к общему количеству материалов, содержащих упоминания об объекте, в процентах.

К сущностным характеристикам внимания (присутствия) должен быть отнесен «знак отношения» со следующей шкалой: негативный, позитивный, нейтральный, смешанный.

Знак отношения источника не следует путать с окраской контекста (повода), в связи с которым проявляется внимание к объекту. Так, заведомо негативную эмоциональную окраску имеет или вызывает контекст, связанный с бедствиями и т.п., но упоминание объекта здесь может быть позитивным (на этом возможно строить продвижение объекта: он/оно/она присутствует там, где нужна помощь).

Данные анализа контекстов (скорее всего, полученные не автоматизированным способом), могут формироваться по специальному запросу клиента. Единичные показатели, характеризующие отдельные тексты, сосуществуют с множественными показателями, характеризующими совокупность текстов; последние характеризуют свойства поведения каналов/источников или положение объекта в медийном поле.

Множественные показатели могут быть «привязаны» к различному количеству каналов, к различному количеству объектов, к различным пространственно-временным рамкам: в одном номере/выпуске, в одном канале, за различные единицы времени, в различном объеме, измеряемом в строках, в квадратах, в иных мерах площади.

Индекс Интегрум

Специалисты по массовым коммуникациям и журналистике, по рекламе и PR все чаще говорят о рынке внимания, как о том пространстве, где каналы информации, с одной стороны, и аудитория — с другой, проявляют избирательное отношение к тому, что происходит вокруг, к тем, кто в этом участвует. Как и на всяком рынке, здесь есть свой капитал (известность), конкуренция и своя «валюта» — рейтинги каналов.

Компания «Интегрум» разработала совместно с профессором кафедры социологии журналистики МГУ И.Д. Фомичевой новые инструменты для описания и анализа рынка внимания, которые имеют общее название «ИндЭкс Интегрум» (IndEx Integrum). Индекс экспонирования «Интегрум» связывает меру внимания канала информации к объекту с величиной его потенциальной аудитории (рейтингом).

Понятие «экспонирование» («exposure») означает контакт с каналом/источником как первую фазу процесса взаимодействия аудитории с сообщением. Чем выше значение ИндЭкс «Интегрум», тем больше внимания получил объект. Индекс вычисляется как произведение количества упоминаний/ публикаций об объекте в выборке и размера средней аудитории номера/AIR теле- или радиопередачи, посещаемости сайта канала/совокупности каналов информации.

¹ Сюжетная (суб)единица — фрагмент материала в теле- и радиотрансляции, ограниченный сменой персонажей, способа представления (запись, студийный текст, архив и т.п.), времени и места происходящего.

Индекс «Интегрум» по смыслу близок к показателю, применяемому в медиапланировании, «количество возможностей увидеть» (Opportunity To See, OTS). Для запоминания объекта важны любые его упоминания, поэтому Индекс экспонирования включает все их виды, является готовым инструментом для оценки положения объекта на рынке внимания.

$$\text{Индекс} = \frac{\text{количество упоминаний объекта в отдельном канале}}{\text{размер аудитории канала}}$$

$$\text{Гросс Индекс} = \frac{\text{общее количество упоминаний объекта во всех каналах}}{\text{размер совокупной аудитории всех каналов}}$$

В систему индексов «Интегрум» входят производные индексы, основанные на базовом индексе экспонирования. Применяя их, клиенты Интегрума могут решать разные задачи. Так, для учета характера экспонирования объекта можно выделить среди всех упоминаний только те, которые формируют положительный его образ. В этом случае применяется Индекс-позитив Интегрум, который равен процентному отношению позитивных упоминаний к общему числу упоминаний объекта:

$$\text{Индекс-позитив} = \frac{\text{Сумма позитивных упоминаний объекта в канале/ах}}{\text{Общая сумма всех упоминаний объекта в канале/ах}} \times 100 \%$$

Положение объекта на рынке внимания мы предлагаем определять как «долю на рынке внимания СМИ», сравнивая значение Индекса экспонирования для каждого из однопорядковых конкурирующих объектов:

$$\text{Доля рынка внимания (ДРВ)} = \frac{\text{Индекс объекта}}{\text{Сумма Индексов всех конкурирующих объектов}} \times 100 \%$$

Работая с системой Индекс Интегрум, можно оценить и долю на рынке позитивного внимания:

$$\text{Доля рынка позитивного внимания} = \frac{\text{Индекс-позитив данного объекта}}{\text{Сумма Индекс-позитивов всех конкурирующих объектов}} \times 100 \%$$

Система индексов Интегрум нацелена на дальнейшее собственное развитие с учетом запросов клиентов.

Комплексное качественное исследование в русскоязычной блогосреде

Шамыкина Виктория Михайловна,
Федеральная служба охраны, г. Москва

Естественная трансформация социальной реальности на текущем этапе развития российского общества вполне явно включает в себя процессы, обуславливаемые «сетевой жизнью» членов этого общества — активностью в среде Интернет. Каналом «смешивания» социокультурных процессов, паттернов поведения, ценностных аспектов, является пользователь глобальной сети.

Формирование в России информационного общества по оценкам некоторых специалистов, идёт кризисным путём. Как минимум, оно проходит с нарастанием скоростей, при котором прогнозирование становится всё более и более затруднительным, из-за возрастания нелинейности траекторий развития. Кроме того, многие исследователи обеспокоены ситуациями нестабильности в сфере духовной жизни, культуры, образования. В числе первых выделяются такие процессы, как развитие разнообразных девиаций, нарушения процессов социализации, развитие различных, не существовавших ранее форм социальной депривации, увеличение уровня агрессии, Интернет-аддикция (зависимость). При этом признаётся, что новые формы коммуникации несут противоречивый характер: указанные проблемы являются лишь одной стороной происходящих в Интернет процессов. Есть и позитивные социаль-

ные проявления использования артефактов этой среды. Рационализация быта, успех коммуникации при географической разбросанности респондентов сообщений, наличие удобных средств для ведения совместных творческих проектов, способных преумножить культурный капитал групп и обществ — вот лишь часть позитивных возможностей использования среды. Сама по себе среда — это комплекс инструментов, и является лишь проводником и площадкой для тех или иных действий.

Феномен Интернет заключается в его открытости, публичности, трудно представимой вне данной среды. Подавляющие объёмы информации, публикуемой в Интернет, могут быть в любой момент доступны произвольному пользователю — как минимум, для ознакомления, а обычно также для комментирования. Именно благодаря такой публичности процессы в среде обладают чрезвычайной динамичностью, и способны оказывать мгновенное влияние на социокультурное поведение как участников среды, так и их ближайших «коллег» по коммуникации, получающих информацию изустно («в Интернет говорят, что...»).

С учётом мощи и динамизма этого уникального инструмента коммуникации, становится очевидной необходимость исследовать его социальный потенциал для российского общества. Материал, на котором возможно такое исследование, уже имеет достаточные объёмы — большая часть сервисов, изначально разрабатывавшихся в Европе и Америке, благополучно используется русскоязычными пользователями, или имеет русскоязычные аналоги (поисковые машины¹, службы хостинга², электронная почта, социальные сети, блоги³). При этом материал не только широк, он имеет уже определённый «срок давности», временную протяжённость, благодаря которой возможно изучение развития определённых процессов и их элементов в динамике. Объёмы данных достигли того уровня накопления, при котором расширение поля исследования грозит переизбытком социологической информации, которую нельзя будет обработать в разумные сроки.

Наша статья ставит своей целью рассмотреть особенности социологического исследования в такой части социокультурного коммуникативного пространства Интернет, как русскоязычные блоги. Выбор именно этой грани социальной жизни в Интернет обусловлен особенностями блогов, позволяющими им отражать практически все социальные процессы общества. И это не просто отражение, отображение. Политическая нестабильность, вопросы образования, медицины, культуры, нравственности, социальные проблемы — феномены как ежедневных обыденных практик, так и практик среды Интернет — всё это находит здесь и активное развитие, трансформацию. Этой тесной связью с общей социальной реальностью и интересна блогосреда для социологического исследования.

Блогосреда, как объект социологического исследования, обладает определёнными качествами, важность которых необходимо учитывать. Понимание особенностей того поля, на котором предстоит работать, складывается из начальной подготовки и личных наблюдений. Под начальной подготовкой мы понимаем, в первую очередь, умение исследователя пользоваться артефактами среды и понимать их семантическую нагрузку. Сюда включено как базовое умение пользоваться компьютером (набирать текст, пользоваться браузером, работать с интерактивными Интернет-страницами), так и ориентироваться в явлениях собственно среды: нетикет, активные на данный момент словари сленга, символические эмоциональные маркеры («смайлы»), общепринятые нормы публикуемой информации и так далее.

Ниже будут подробно рассмотрены особенности и тенденции среды, оказывающие влияние на поведение участников. Однако, в рамках обоснования подхода, суть которого излагается в статье, необходимо обобщить некоторые черты русскоязычной блогосреды. Она объе-

¹ Яндекс, Рамблер — как правило, сайты этих поисковых машин имеют массу дополнительных сервисов: возможность бесплатной регистрации адреса электронной почты, сервис электронных досок объявлений, поиска вакансий и т.п.

² От англ. «host» — «общий термин, описывающий нечто, содержащее ресурс и предоставляющее к нему доступ» (из словаря Lingvo), чаще всего употребляется для обозначения сервиса платной или свободной аренды пространства на жестких дисках, для публикации т.и. данных в Интернет.

³ Сервис, предназначенный для регистрации и ведения публичных «сетевых дневников», будет рассмотрен подробно ниже.

диняет в себе несколько сервисов, предоставляющих сходные по сути услуги: возможность создания сетевого дневника, блога – страницы, устроенной определённым образом, а также ряда страниц, связанных с основной. Это всегда страница профиля, где содержится та информация о пользователе, которую он решил опубликовать, а также ряд опциональных страниц, устройство которых разнится от сервиса к сервису. Суть же страницы самого дневника в том, что записи, которые в нём могут быть созданы, датируются и расставляются в порядке убывания даты: от самых новых к более ранним. Но дневниковый характер не ограничивается лишь расположением записей; как правило, он присутствует и в контенте, т.е. содержанием, и отвечает общему представлению о том, что такое «дневник» с поправкой на Интернет-реалии.

Сетевые дневники или блоги представляют собой фрагменты медиа-пространства, объединённые как по признаку принадлежности тому или иному сервису (каждый из которых обладает своей спецификой), так и по типам (блоги новостные, личные и проч.). По своей сути все технические артефакты блогов являются комплексом Интернет-сервисов, не представляющих никакой новизны. Блог похож на личную страницу, ставшую его прообразом. Блог похож на форум в своей дискуссионной части. Сообщества, где преобладает коллективное авторство, еще ближе к форумам. Подавляющее большинство современных сервисов предоставляет службу внутренней почты — это полный аналог электронной почты, ограниченный пределами сервиса. Расположение записей в хронологическом порядке делают возможной публикацию новостей, которые тут же доступны как широкой аудитории (если доступ на блог не ограничен), так и подписанным на блог участникам сервиса. Такое сочетание сервисных инструментов и аккумуляция информации в определенных фрагментах среды делает блоги уникальным явлением. Кроме того, в рамках сервиса его пользователи оказываются в группе блоггеров, приобретая специфическую групповую самоидентификацию, которая не мешает дифференциации внутри этой группы.

Артефакты среды можно отнести к двум функциональным группам: это связеобразующие инструменты и содержательные площадки. Первая группа отвечает за создание и оформление разнообразных ссылок, связывающих пользователей блогосреды, тем самым создавая более или менее тесные общности в рамках одного сервиса или нескольких. Содержательные площадки — это отчасти те же артефакты (например, комментарии, которые отображают и связи, и контент), изучаемые в аспекте информационной, смысловой, семиотической наполненности.

Исходя из особенностей среды, учитывая нелинейность и сложность процессов, мы рассматриваем качественный подход как основу исследования. Этот выбор обусловлен тем, что блогосреда представляет собой пересечение множества личных пространств, и в аспекте взаимоотношений, социализации и ресоциализации участников, усвоения ими ежедневных практик наиболее динамична. Количественные подходы здесь обладают не достаточной информативностью при рассмотрении символов, знаков, смыслов среды, когда как качественное исследование помогает верно их интерпретировать, и за счёт этого исследовать феномены, оцениваемые как негативно (аддикция, агрессия, усвоение отклоняющегося поведения), так и позитивно (творческая активность, взаимопомощь, креативность).

Выбор метода социологического исследования социальных процессов в блогосреде требует представления о репрезентативности. На какую социальную группу могут быть распространены полученные выводы, может ли средний русскоязычный автор сетевого дневника быть приравнен к среднему россиянину? В России включение в эту среду характерно, в первую очередь, для молодежи 15–25 лет, и основную долю среди этой совокупности имеют жители крупных городов: Москвы, Петербурга, Екатеринбурга, Краснодара и др. Эта проблема представительности, в частности, раскрывается в работе Л.Л. Делицина¹. Представленные им результаты исследования уровня проникновения Интернет в России указывают, в первую очередь, на важность работы со школьниками, то есть, подключение к сети школ со-

¹ Делицин Л.Л. Скоро ли пользователь Интернета станет похож на среднего россиянина? // Онлайн исследования в России 2.0 / Под ред. А.В. Шашкина, И.Ф. Девятко, С.Г. Давыдова. М.: РИЦ «Северо-Восток», 2010.

гласно национальной программе «Образование». По представленным в работе прогнозам основу пользователей и дальше будет составлять молодежь. Важность исследования обусловлена тем, что актуальные модели поведения создаются и воссоздаются (а) для перечисленной выборки в среде Интернет и сетевых дневников и (б) в относительной изоляции от моделей старшего поколения (которое исключительно слабо представлено в сетевых дневниках, в основном, это знаменитости и известные творческие личности).

Общие этапы предлагаемого комплексного подхода таковы. Процесс «вхождения в поле». Здесь возможны два подхода. Первый: самостоятельное знакомство, при котором требуется длительный период наблюдения именно среды, её устройства и обыденных практик, открытого общения с участниками среды. Самостоятельная расшифровка символов и знаков среды возможна с помощью уже накопленных знаний (сайт lurkmore.ru как пример «энциклопедии современных смыслов») и при личном изучении употребления тех или иных семантических и семиотических «приёмов». Второй: с проводником, опытным пользователем сетевых дневников. В этом случае процесс вхождения ускоряется, но широкий массив неиспользуемых проводником символов и знаков оказывается опущен. Таким образом, рекомендуется использовать оба подхода: имитацию действий новичка в среде (это, например, позволит задавать уточняющие вопросы) и общение с проводником для исследования знакомого ему фрагмента среды, как одного из множества устроенных аналогично.

Первичное наблюдение начинается со знакомства со средой, выбора случаев (акций интересующих исследователя феноменов) и авторов. Для этого может быть удобно использовать инструмент «случайного дневника» и такие инструменты среды для отбора интересующих случаев как статистика, поиск, разнообразные списки авторов и сообществ. Набор «опыта» при простом неструктурированном наблюдении позволит более четко сформулировать исследовательские задачи. Здесь же важно выбрать, будет ли наблюдение скрытым или открытым, и результирующий подход зависит от этических воззрений исследователя.

Следующим этапом является оценка информативности выбранного сетевого дневника. Здесь есть «подводные камни», некоторые сложности в оценке информативности дневника при учёте исследовательских задач. Так, некоторые современные исследователи, например, В.И. Ильин, говорят о малой информативности «гулких комнат», блогов с низкой активностью¹. Это утверждение кажется нам достаточно спорным: при исследовании обыденных моделей поведения, например, такие блоги могут предоставить принципиальную информацию.

Четвертый этап — работа со смыслами и группами. Язык есть способ разграничения групп. Анализ смысловых единиц и выражений позволяет определить общие тенденции среды (распад и дробление единиц «олбанского языка» в качестве примера). Границы групп представляют собой пределы «понимания» тех или иных форм языка. Анализ связей используемого автором языка, группы, к которой причисляет себя автор, и моделей поведения, реализуемых автором, ведение включенных в карточку наблюдения «языковых профилей» авторов, наблюдаемых подробно, позволяет воссоздать социологический портрет субъекта блогосреды. Психолингвистический подход к изучению языка блогосреды помогает исследователю при анализе не только семантико-семиотических аспектов, но и позволяет упорядочивать набранные языковые единицы по различным основаниям. Так, при исследовании активных блогов сервиса @diary.ru, был сделан вывод: популярность является формой обратного влияния аудитории на автора. Авторы популярных блогов пишут специфическим языком, сочетающим богатство словаря и разнообразие выразительных средств, с одной стороны, и сленговые единицы с экспрессивной окраской — с другой².

Структура сетевого дневника и блогосреды представляет собой своеобразный аналог «документов жизни», аналог этнографической площадки. Активные дискуссии и работа сообществ — это аналог «фокус-групп», скрытое включённое «фокусирование», при котором ис-

¹ Ильин В.И. Драматургия качественного полевого исследования. СПб.: Интерсоцис, 2006.

² Активными были признаны блоги по таким количественным показателям, как частота обновления и число посетителей, а также по качественным показателям, таким, как характер отношений между автором и его посетителями, темы, используемые автором, и его мировоззренческие, ценностные характеристики.

следователь задает тему дискуссии и направляет ее на правах одного из авторов-участников среды. Здесь можно также применить метод глубинного интервью с наиболее информативными авторами дневников, являющихся площадками активных коммуникативных процессов, обмена и генерации моделей поведения. Преимущества и «аналогов» исследовательского материала, предоставляемого средой: электронный вид, ссылочная структура, цитирование и доступ к архивам, с одной стороны. К недостаткам можно отнести ограничение доступа, сложность включения в различные группы, многообразие смыслов, трудность расшифровки и избыточность материала.

Комплексность такого обобщенного подхода заключается в следующем. Во-первых, исследовательская позиция относительно блогосреды, как фрагмента медийных сред Интернет, предполагает не только взаимодействие с этим фрагментом, но и более широкое представление о сетевом общении. Блогосреда не является ни изолированной, ни просто отражающей все процессы, но представляет собой совокупность артефактов, формирующих поле, дающее обширную информацию о современных социокультурных процессах, особенно, в среде молодёжи. Н.Л. Соколова и Л.А. Василенко подчёркивают важность исследования этой среды, сложной за счёт возрастающей нелинейности, свойственной динамике развития¹. Подобные среды комплексны по своей природе.

Во-вторых, исследователь, имея дело с техническими артефактами, должен помнить, что, на основании жестко запрограммированных инструментов, предлагаемых сервисами блогов, индивиды проявляют творческую и свободную активность, что позволяет сочетать формализованные и неформализованные подходы. Аналитики, обращающиеся к медийным средам Интернет, в том числе, к блогам, подчеркивают их неформализуемые грани: отношения мнения, тональность высказываний и т.д. В этом случае сочетание количественного анализа с качественным оказывается порой весьма эффективно.

Наконец, позиция исследователя относительно сложных процессов среды может быть динамична, отвечая изменчивости самой среды. Указанные выше аналоги социологической информации требуют каждый раз нового осмысления. Исследование дневников, как документов жизни, задействует массив информации способом, отличным от этнографического подхода к этому же массиву. Создание фокус-группы на базе сетевого сообщества имеет свою специфику за счёт особенностей блогосреды. Наблюдение, в повседневности всегда осуществляемое в реальном времени, в блогосреде может быть ретроспективным, за счет того, что среда сохраняет большую часть данных о предыдущих состояниях. В случае, когда интервьюирование по электронной почте дополняется личным, могут быть не только получены уточняющие данные, но и дополнены интерпретациями символы и знаки, используемые в сетевом общении.

Предлагаемый в статье подход, в первую очередь, акцентирован на изучении максимального числа источников, и изучении максимально же многообразном. Комплексный подход позволяет ознакомиться с широким полем проявлений коммуникации в среде сетевых дневников, и затем концентрироваться на тех способах и источниках получения данных, которые соответствуют исследовательской задаче. Пилотажные исследования с использованием предлагаемого подхода показывают, что таким образом представление о сложной структуре среды, состоящей из множества системных единиц, достигается успешнее, чем при изучении лишь части феноменов. Такой метод, усложняющийся с расширением числа включенных в него частных подходов, требует дальнейшей разработки, поскольку и среда усложняется с течением времени. Апробация предлагаемого подхода и его различных частей и элементов, их развитие необходимы, поскольку исследование медийных сред является задачей актуальной и значимой в современном развивающемся обществе.

¹ Соколова Л.Н. Популярная культура Web 2.0: к картографии современного медиаландшафта. Самара: Самарский университет, 2009; Василенко Л.А. Интернет в информатизации государственной службы России (социологический анализ). М.: Наука, 2008.