

*С.Г. ПАШКОВ*

### **СЕМАНТИЧЕСКИЙ СЕТЕВОЙ ПОДХОД: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ (ПРИМЕР ОБРАЗА ИНФЛЯЦИИ В СМИ)**

*Аннотация.* В статье акцентируется потребность в инструментарии, который позволил бы облегчить построение системы кодирования и анализа информации новостных сообщений. Материалом для исследования послужили экономические новости, изобилующие специфическими терминами, интерпретациями, экспертизой и метафоричным описанием событий. Во многих случаях можно фиксировать, что содержание подобных текстов усложнено, в результате чего «классический» контент-анализ может потребовать дополнительных итераций и усиленного внимания к аналитической процедуре. В исследовании освещаются методологические, аналитические особенности семантического сетевого подхода (ССП) в сравнении с контент-анализом и подходами Text Mining на примере анализа шести новостных сообщений экономической тематики, содержащих термины «рост цен» и «инфляция». Особенность СПП — упрощение обработки больших неструктурированных данных с акцентом на содержании. В подготовке и расчете сетевых метрик для каждой новости отражаются наиболее значимые концепты, что упрощает контент-анализ большого корпуса текстов. Визуализация показывает в кейсах разное семантическое положение «инфляции» как синонима «роста цен» в зависимости от сюжета. Важным результатом является то, что вне зависимости от объема и визуальной структуры новостного сообщения эти термины можно считать ведущими в соответствующих сюжетных линиях, что может помочь в проведении дискурс-анализа с их упоминанием. Предполагается, что подход станет «опорным» инструментом для дальнейшего количественного и качественного анализа новостных сообщений, в частности с экономической тематикой. Технические особенности программ подготовки текстов и семантического

---

**Пашков Станислав Георгиевич** — аспирант, преподаватель факультета социальных наук; стажер-исследователь Лаборатории экономико-социологических исследований, Национальный исследовательский университет «Высшая школа экономики».

**Адрес:** 101000, Москва, ул. Мясницкая, 11. **Телефон:** +7 (495) 772-9590\*12450.

**Электронная почта:** spashkov@hse.ru

моделирования можно считать возможными ограничениями подхода, особенно для пространства Text Mining.

*Ключевые слова:* социально-экономическая информация; новостной текст; семантический сетевой подход; семантические карты; процессинг текстов; контент-анализ; дискурс.

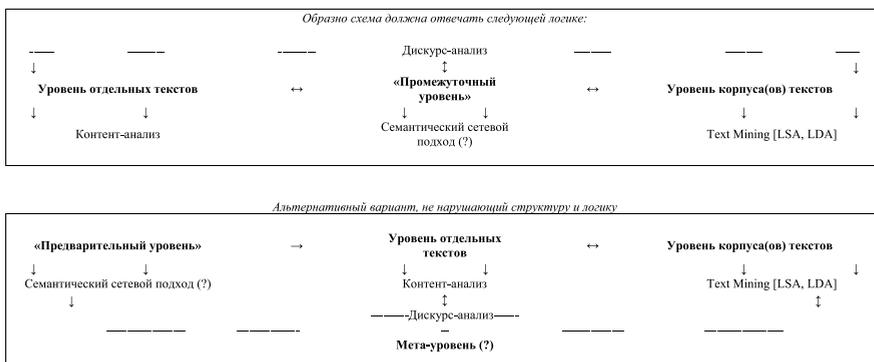
*Для цитирования:* Пашков С.Г. Семантический сетевой подход: возможности и ограничения (пример образа инфляции в СМИ) // Социологический журнал. 2020. Том 26. № 2. С. 8–30. DOI: 10.19181/socjour.2020.26.2.7262

## **Введение**

В социологии массовой коммуникации наблюдается движение в сторону новой методологической парадигмы «интеллектуального анализа текстов» (Text Mining). Остаются ли «классические» методы анализа (контент-анализ [4], дискурс-анализ) валидными, если при росте объемов данных не упрощается анализ их структуры? Этот вопрос актуален в сфере изучения новостных сюжетов (*экономических*), где преобладают сложные термины, апелляции к показателям [7], усложняются описания экономической реальности, что может влиять на экономические стратегии населения. Проблема в том, что при необходимости совмещения числа, тональности, тематики и внутреннего содержания сообщений необходим пересмотр логики исследования. Например, когда важно понять, как события в СМИ «работают» на экономические решения населения, возможен ли анализ смыслов в купе с количественными подтверждениями? И каким образом это можно сделать без дополнительных временных издержек? Предлагается к обсуждению *семантический сетевой подход* (ССП) и его возможная связь с контент-анализом и автоматизацией обработки текстов. Цель статьи — выделить особенности ССП, его преимущества и недостатки, сравнивая с некоторыми существующими подходами анализа текстов. Вначале показана реконструкция ССП, затем работа этого подхода на примере серии текстов, а в конце подводятся итоги, делаются содержательные выводы относительно применимости подхода.

## **Место ССП в стратегиях анализа текстов**

Актуальные социальные исследования демонстрируют практики контент-анализа, дискурс-анализа, алгоритмов Text Mining. Однако возникает вопрос, возможно ли совмещать эти методы при анализе больших корпусов текстов, что при текущих потоках данных, несомненно, и практично и интересно [17]. Может возникнуть потребность в количественных данных, но при отборе кейсов потребуются внимательный анализ самого содержания. Предлагаемый к обсуждению подход может стать важной «прокси-компонентой». Однако при каких условиях следует применять именно его? Рассмотрим рисунок, чтобы выделить позицию ССП среди других логик работы с текстом, а также основные затруднения, связанные с каждой из них.



**Рис. Предполагаемое место ССП в системе методов анализа корпусов текстов**

В *контент-анализе* есть принцип, направленный на оценку факта присутствия (выраженности) и «работы» различных аналитических категорий в тексте, включая сравнение и сопоставление коммуникационных моделей, описание события либо явления [21]. «Качественная» сторона анализа предполагает понимание структуры текстовых данных. В контент-анализе можно выделить ряд направлений. Это, например, «прагматический контент-анализ», с фокусом на частоте наблюдаемых знаков; «семантический контент-анализ» — классификация знаков в соответствии с их содержанием, упоминанием определенных лиц, групп и др. [21, с. 45]. Несмотря на разнообразие предоставляемых возможностей, метод достаточно ресурсоемкий. Исследователю для извлечения максимума из материала необходимо заранее выработать набор рабочих вопросов и изучаемых (проверяемых) категорий, а также инструментарий для их фиксации и анализа. По ряду научных разработок можно заметить наличие: 1) высоких издержек *ручного* сбора данных, уточнения выборки; 2) требований тщательной подготовки исследователя к выделению рабочих концептов; 3) акцентов на специфике содержательной интерпретации, трудностей «перевода» ряда категорий в социальные факты. Конечно, сегодня существуют специальные пакеты, через которые это можно сделать (к примеру, анализ сентиментальности сообщений с помощью ATLAS.Ti [17]). Но кодирование исследователем либо кодировщиками все равно подразумевает «медленный» и *вдумчивый* анализ текстового контента, что особенно важно для качественного подхода [12]. Более того, экспоненциальный рост числа публикаций в средствах массовой информации, электронных новостных агрегаторах приводит к естественным сложностям в построении оптимальной выборки и ее обработки<sup>1</sup>.

<sup>1</sup> Контент-анализ значимо связан в этом вопросе с фрейм-анализом, речь о котором далее.

Решить эти проблемы может парадигма *Text Mining*, где доминирующую роль играют алгоритмы обработки данных большой размерности, теория вероятности и программирование. Основные концептуальные и эмпирические разработки описаны Дж. Эвансом и П. Ачевесом [16]. *Text Mining* изначально ориентирован на обработку любых данных с помощью ЭВМ, с трансформацией логики представления текстовой информации. В случае с контент-анализом мы имеем дело с разными типами «единиц» (units), включая целостный текст, абзац или отдельное слово. В *Text Mining* объектом является документ как целостный конструкт, внутри которого просматривается специфический вектор слов. Само по себе содержание документа становится как бы чуть менее «значимым», нежели его объем и внутренняя структура. Данный подход исключительно количественный, благодаря чему есть возможность работать с большими корпусами текстов, охватывающими тысячи и более текстов на разных языках. Примером в социологии может быть исследование П. ДиМаджио и соавторов [13], а в сфере массовых коммуникаций — работа К. Якоби и соавторов [20]. Их объединяет применение «скрытого размещения Дирихле»<sup>2</sup> (*Latent Dirichlet allocation, LDA*) — техники тематического моделирования текстов с помощью машинного обучения, хотя и здесь есть некоторые сложности. Например, требуется подготовка текстов: нужно не только задать математические параметры для алгоритма, но и провести достаточно тщательную разметку текстов (исключение знаков препинания, союзов и другие лемматические операции). Также нужно уделять внимание и выбору алгоритма, поскольку их довольно много, что потребует понимания технических аспектов процессинга текстов. В итоге сам анализ становится комплексным, внимание исследователя смещается на «погружение» в контекст отдельно взятой статьи, что неявно демонстрируется в исследованиях.

В качестве дополнения можно обратить внимание на *дискурс-анализ* [30]. Он находится над ССП в нашей схеме потому, что основной акцент делается на осмыслении и анализе речевых конструкций и смыслов в текстах (например, посвященных экономическим событиям), что может использоваться в контент-анализе и, возможно, в *Text Mining*. Фактически метод полезен тем, что позволяет исследователю задавать качественные «рамки» анализа противоречий внутри отдельно взятого текста, определять социальную значимость рабочих и вторичных концептов. Иными словами, выполнять процедуру аналитической

---

<sup>2</sup> Алгоритм реализует принцип, при котором весь текстовый корпус рассматривается в качестве своеобразного «мешка слов» (Bag of Words), с разной частотой одних и тех же слов в качестве параметров. С помощью вероятности оценивается корреляция конкретного экземпляра текста со списком «тем» (частот наиболее употребляемых слов). LDA наиболее «устойчив» в плане интерпретации [15] и как вариант количественного контент-анализа текстов с последующим получением статистически надежных результатов.

валидации концептов перед количественным анализом их сетевой взаимосвязи. А сам семантический сетевой подход может занимать «среднее» положение именно потому, что он концептуально вбирает в себя и качественную, и статистическую составляющие анализа текстов. И здесь уже необходимо разобраться в его архитектуре.

### Основания семантического сетевого подхода

Исторические корни подхода лежат в компьютерных и социальных науках. Центральный компонент — *семантические сети* (раннее название, данное основателем направления Р. Квилианом в 1960-х, — семантические карты) — отражают содержательные отношения слов, подкрепляя это математическим аппаратом анализа. Развитие направления было весьма противоречивым в эпоху слабых ЭВМ 1960-х гг. [9], однако впоследствии были определены полезные приложения, позволившие решить некоторые проблемы. В частности, приобретают свое развитие исследования структуры содержания массовой информации в рамках теории повесток дня третьего уровня<sup>3</sup> [19]: социальные сети (графы) используются для изучения отношений между характеристиками массовой информации и ее представлением в форме сообщений, например понимания того, «сохраняет ли аудитория взаимосвязь между [контекстами] в [массовых] сообщениях, если не [требуется] углубленное обоснование» [19, р. 617].

Можно выделить концептуальное и исследовательское поля ССП. Первое хорошо обозначено в работах Дж. Соуы [27], К. Карли [12] и др. Семантические сети представляют собой систему концептов («точек»), объединенных направленными связями («стрелками»). Эта система не располагает заранее заданным паттерном и является уникальной для отдельно взятого текстового массива. Ф. Лиман отмечает, что она хранит не только «фиксируемую сетевую структуру, но также и методы для автоматического извлечения из нее гораздо большей структуры, совокупности подразумеваемых знаний» [22, р. 2]. В исследовательском поле программа кодирования текстов составляется с учетом выбранных ключевых категорий, заканчивая уровнем генерализации. ССП оперирует принципами расчета сетевых показателей и семантическим картированием (преимуществом является интерпретация схем категоризации), сведением сложных текстовых явлений при условии минимальных представлений об отношениях слов в тексте.

---

<sup>3</sup> Для понимания логики теории повесток дня кратко опишем основные идеи двух ее первых уровней. Разработкой данной концепции занимались в основном С. МакКомбс и Р. Шоу. Согласно качественному обзору А. Казакова [1], первые два уровня отражают стратегию освещения тем и атрибутов (свойств) их объектов, соответственно обозначая их в качестве первого и второго уровней воздействия на массовое поведение, которое также обозначается как *salience*. «Типичная» повестка дня характеризуется сюжетной линией, позицией и тональностью сообщения.

### Методическая составляющая ССП

Далее, важны примеры реализации подхода. В книге С. Симоффа и соавторов [25] представлено разнообразие методологических приемов семантического анализа (в т. ч. через сетевой анализ) и демонстрируются формы визуального текстового анализа из области «Майнинга данных» (Data Mining) с помощью вычислительных алгоритмов. Для социологии такие подходы являются комплексными; необходим более простой методический принцип обработки и визуализации текстовых данных. Поэтому в контексте ССП большую ценность имеют работы Ф. Дрейгера [14] и С. Эддингтона [15]. Основываясь на результатах исследований, мы полагаем, что в социологической плоскости ССП проработан недостаточно. Чаще он рассматривается как подход, обособленный от других методик исследования текстов. В таком случае исследователь выбирает между глубиной анализа и объемом материала. В настоящем исследовании прослеживается, как концептуализация объекта берется у контент-анализа (по большей части «качественного»), поскольку необходимо ставить и решать *содержательные* цели и задачи. А из Text Mining заимствуется логика обработки и представления количественных данных. В таблице 1 представлена эта логика.

Важно отметить, что представленную в таблице 1 последовательность можно рассматривать как *универсальную*, подходящую для исследователей с разным бэкграундом в компьютерной обработке информации. В некоторых случаях бóльшая эффективность достигается с помощью специальных программных пакетов, написанных для языка R.

Однако есть проблема: заранее подготовленной категориальной системе необходима операционализация понятий либо слов. Чаще всего установить такую систему оказывается непросто без четко формулируемых вопросов, обстоятельно проработанной системы концептов, их интерпретации и связи с предположениями (гипотезами). Более того, сложные термины и числа могут нарушать устойчивость системы категорий, вследствие чего при выходе на уровень генерализации полученные результаты могут быть неопределенными. Для этого случая исследователю необходим более формализованный, стандартизованный и наглядный подход, коим может стать ССП. Безусловно, результаты нашего исследования следует рассматривать как пробные, поскольку они получены на сравнительно небольшой выборке текстов разных периодов. Мы анализируем разные случаи описания событий с применением экономической информации на примере категорий «инфляция» и «рост цен», рассматриваемых в качестве базовых. Другие концепты (например «безработица», «кризис», вариации экономической информации) рассматриваются в качестве дополнительных и вторичных. С помощью ССП мы ищем качественный и метрически обоснованный «минимум» концептов, необходимую оценку содержательной устойчивости того или иного новостного текста.

Таблица 1

### Процедурная схема проведения семантического сетевого анализа с использованием прикладных пакетов MyStem, AutoMap, ORA и R

Шаг	Наименование шага	Описание / Характеристика	Софт
№ 1	Сбор текстового корпуса	Поиск и накопление массива данных, содержащего новостной материал. «Вычитка» текста на соответствие теме и объекту(-ам) исследования	Web-браузер Редакторы
№ 2	Первичная предобработка	Сведение и очистка «дубликатов», исключение знаков препинания и экономических, математических знаков	Редакторы AutoMap
№ 3	Финальная предобработка текстового массива	Проведение серии процессов: лемматизация (приведение слов к начальной форме), стемминг (сокращение слов без окончаний), токенизация (формирование текстового корпуса как «мешка слов», с разделением на предложения)	MyStem AutoMap [Редакторы]
№ 4	Генерация семантической сети	Определение наличия / отсутствия требований по направлению семантических отношений (bi-directional либо uni-directional). Использование настроек «Generate Semantic Network (DyNetML)» <sup>3</sup> , выбор «ширины семантической связи» (число слов, которые связаны семантическими отношениями) <sup>4</sup>	AutoMap
№ 5	Количественные метрики, анализ	С помощью пакета осуществляется сегментирование семантической сети на смысловые группы. Получаем расчет количественных характеристик [23], включая: 1) Топологическая структура сети: частоты встречаемости концептов, связей и плотности сети 2) Степень центральности (связей конкретного слова с соседними) и посредничества (слова-«посредники» в пространстве смыслов) 3) Центральность собственного вектора и вариации (концепты, находящиеся в одной группе значимых и играющие роль «влиятельных» смысловых единиц)	ORA-Lite либо R
№ 6	Визуализация структуры и ее интерпретация	Визуальная оценка качества формирования микрофреймов в семантической сети (популяции концептов) Определение того, насколько похожими оказываются концепты между собой по смыслу в пределах текста	ORA-Lite либо R

<sup>4</sup> Выбор ширины окна похож на отбор категорий в контент-анализе: это достигается аналитическим путем исходя из представлений исследователя о связях между концептами. Для всех текстов в нашей выборке подошло «усредненное» значение 3. С точки зрения расчетов количество связей при увеличении ширины семантической связи растет; в то же время сетевая плотность может снижаться. Избыточность связей может исказить качественный анализ сетевых отношений концептов.

### Концептуализация экономических новостей и их анализа

Поскольку тема влияния СМИ на поведение людей хорошо изучена и в данной области выработано много подходов [29], необходимо проводить отбор текстов по какому-либо критерию. Для ССП имеет смысл рассматривать подходы, связанные с фреймированием новостных текстов как аналитическим представлением семантических карт, но могут быть применены и другие подходы.

Теория фреймов находит выражение во многих эмпирических направлениях. Существует точка зрения, что трактовки фрейминга различны еще и в силу взаимосвязи с теорией повесток дня. Тем не менее возникает основной вопрос — как информация доходит до массового потребителя и в какой форме «огрубляется», фиксируется в качестве темы [11]. Проблема «многополярности» трактовок, если подразумевать вопрос выбора рабочих интерпретаций, остается при этом значимой. Text Mining релевантно «определяет» фреймы как наборы очерченных совокупностей текстов, включающие ассоциативные стереотипы, подразумеваемые в структуре данных, где буквально одно слово представляет «категорию опыта». Конечно, такая трактовка своеобразна, т. к. М. Минский (автор теории фреймов) рассматривал способы репрезентации знаний и ассоциаций в памяти человека, чтобы позднее перенести их в компьютерные алгоритмы [9]. Это можно представить как упаковку новостных сообщений в отдельные рамки (фреймы), вместе со статистической оценкой качества этой упаковки алгоритмами машинного обучения. Трактовка фреймов Р. Энтманом также имеет важное значение. Под ними понимаются действия по отбору и выделению значимой для людей реальности, которая содержится в некотором «сигнале» внутри сообщения. Эта логика подразумевает, что новостной текст не рассматривается как некоторая текстовая рамка [3]; скорее мы должны обратить внимание на тему новости с ее условными «сигналами», что способствует погружению читателя в контекст (иначе говоря, «фреймирует» его восприятие события). Анализ таких фреймов, конструируемых журналистами и массовыми изданиями (институциональными агентами), возможно сделать с помощью методологии *фрейм-анализа*, который открывает перед исследователем большее число научных разработок и подходов<sup>5</sup> и позволяет решить (в некоторой степени) проблему работы с большим объемом текстов.

<sup>5</sup> Может показаться справедливым вопрос, почему на помещенном выше рисунке не присутствует фрейм-анализ как еще один метод анализа текстов, который может быть наиболее релевантным применительно к данному исследованию. Предполагается, что контент-анализ и дискурс-анализ включают в себя компоненты фрейм-анализа. Это хорошо показано в работе Я. Двора и соавторов [2]. Суть в специфике фреймов. А. Казаков указывает, что фреймы можно рассматривать и как медийные (то есть как конструкты, которые «программируются» СМИ для отражения действи-

Однако к концептуализации экономических новостей можно применить и такую логику: рост числа источников, информационных потоков и сообщений ведет к информационным перегрузкам во многих сферах, особенно экономической; негативные сообщения, достаточно долго «поставляющиеся» населению, могут привести к трансформации «реальной» реальности, которая может серьезно отличаться от конструируемой в СМИ [17]. Модель «провала в знаниях» П. Тиченор указывает на различия в качестве усвоения и понимания происходящего в зависимости от социально-экономического положения людей [28]. При анализе текстов разных изданий следует иметь представление об их целевых аудиториях, которые могут различаться по уровню образования. Экономические новости могут быть «перегружены» не только информационно, но и содержательно.

### **Характеристика новостного корпуса**

Дизайн проведенного нами исследования выполнен в виде *кейс-стади*. Неслучайным образом были выбраны шесть новостных сообщений, каждое из которых затрагивает тематику инфляции. При отборе сообщений мы руководствовались результатами опроса россиян об их инфляционных ожиданиях<sup>6</sup>. Были выделены периоды наиболее негативных и наиболее позитивных изменений. Значимые отклонения в реакциях населения на изменения в ценах наблюдались в конце 2010 и 2014 г. Данные периоды кажутся нам наиболее интересными для того, чтобы опробовать предлагаемый подход. Статьи (список представлен в таблице 2) отбирались таким образом<sup>7</sup>, чтобы в них были одновременно представлены концепты *рост цен* и *инфляция* (за исключением одного кейса, содержавшего только концепт «рост цен»).

---

тельности и некоторого шаблона оценивания), и как индивидуальные (попытка построить описательные конструкты, которые должны оказывать некое воздействие на определенные сегменты аудитории) [2, с. 107]. От этого, к слову, зависит не только то, как исследователю нужно рассматривать новостное сообщение, но и то, как кодировать концепты и смыслы. Поэтому есть возможность через ССП преодолеть часть издержек на кодирование информации для акцентирования внимания на оценке, например, типа фреймирования сообщений и их содержательной интерпретации.

<sup>6</sup> Использован мониторинг «Инфляционные ожидания и потребительские настроения населения», публикуемый Центральным Банком РФ. Репрезентативный опрос населения в возрасте от 18 лет (N ≈ 2000 чел. / волна).

<sup>7</sup> База данных EastView использовалась для загрузки текстов.

Таблица 2

**Список статей для анализа и их объем в словах**

Название статьи (текста)	Источник	Объем
1 «Шок на год»	Ведомости. 2010. № 18 [05.10.2010 03:00]	570
2 «Два знака России»	Ведомости. 2014. № 24 [25.12.2014 01:37]	791
3 «Продукты кормят инфляцию»	Gazeta.ru. 14.12.2010, 15:47	518
4 «Цены растут и без спекулянтов»	Gazeta.ru. 13.12.2014, 21:39	953
5 «Что будет с ценами зимой»	Комсомольская правда. 06.12.2010. № 181	1059
6 «Кто не скачет — тот не рубль»	Комсомольская правда. 06.11.2014. № 45	1179

Совокупной объем корпуса анализа (всех шести статей): 5070 слов (концептов, без предобработки)

**Специфика интерпретации данных**

Возвращаясь к поставленной проблеме, мы должны обратить внимание на экономические новости как на пригодный для ССП объект исследования. Такие сообщения воспринимаются в качестве сигналов, фокусирующих внимание на значимых для населения проблемах в экономике. Почему была выбрана *инфляция*? По данным Левада-Центра, наиболее серьезная проблема в жизни россиян — это постоянная инфляция (61%) [8]. При этом, по данным «ФОМ», более «половины россиян [считали в 2018 г.], что состояние экономики освещается в СМИ необъективно; две трети участников опроса уверены, что [СМИ] обнародуют неполн[ую] информаци[ю]» [9]. Как это можно объяснить? Для этого предлагается рассмотреть структуру текстов и выделить в них «типичные» коды (концепты), которые описывают значимость акцента на проблеме роста цен в новостях.

Формально инфляция — это рост цен на товары и услуги, выражаемый в том, что цены становятся неравнозначны имеющейся цене денежной массы, что в дальнейшем ведет к ухудшению социально-экономических условий для населения. Что делается для «разгрузки» информационного потока в сознании читателя путем представления в разных частях текста самого термина и его интерпретаций как «одного и того же»? Для этого нужно не только произвести расчет частоты упоминания целевых концептов (таких как «рост цен» и «инфляция»), но и «сканировать» текст с целью поиска новых значимых концептов. Сделать это можно на основе количественных показателей и их сравнения на межтекстовом уровне (табл. 3).

Таблица 3

**Количественная характеристика работы ведущих концептов в текстах<sup>8</sup>**

Метрика	Текст № 1	Текст № 2	Текст № 3			
Число концептов (фактическое)	211	315	225			
Число их связей (фактическое)	586	903	562			
Плотность сети, %	2,65	1,83	2,23			
Доля успешной кластеризации, %	50,20	47,90	52,50			
Число устойчивых смысловых групп	293	464	272			
Степень центральности (10 концептов с самым большим числом связей в доле от всего текста)	сентябрь	0,611	инфляция	0,784	инфляция	0,781
	непродовольственный	0,532	2015	0,524	год	0,439
	рост цен	0,505	декабрь	0,462	рост цен	0,368
	инфляция	0,479	становиться	0,424	высокий	0,365
	год	0,478	экономический	0,404	прогноз	0,352
	товар	0,463	рубль	0,393	Россия	0,334
	цена	0,437	неделя	0,386	цена	0,326
	0 [число]	0,418	рост_цен	0,383	продукт	0,326
	ускоряться	0,417	население	0,368	рост	0,309
	подорожать	0,368	спад	0,359	питание	0,278
Степень собственного вектора (наибольшая выраженная важность 10 концептов в сети)	сентябрь	0,447	инфляция	0,488	инфляция	0,730
	непродовольственный	0,423	2015	0,375	год	0,456
	товар	0,379	рост цен	0,335	высокий	0,354
	0 [число]	0,371	становиться	0,318	прогноз	0,353
	ускоряться	0,356	спад	0,307	Россия	0,299
	инфляция	0,337	происходить	0,279	следующий	0,299
	рост цен	0,290	год	0,244	продукт	0,294
	подорожать	0,290	население	0,242	цена	0,263
	год	0,273	декабрь	0,237	питание	0,255
	дорожать	0,247	ускорение	0,235	темп	0,201

*Примечание:* В таблице отражены метрики структурного положения и «работы» концептов в сети связей. Степени центральности выражены в долях от всего текста от 0 до 1. Больше значение — более явно работает концепт.

<sup>8</sup> Данные по Текстам № 4–6 и визуализация Текстов № 1 и № 3 — в Приложении.

Перейдем к рассмотрению кейсов. Отметим, что подробная интерпретация количественных данных опущена, а фокус сделан на визуализации и структуре экономических текстов. На рисунке 1 в Приложении представлена семантическая карта для Текста № 1 (газета «Ведомости», 2010). Можно сразу обратить внимание на то, что концепты «рост цен» и «инфляция» в данной статье присутствуют и имеют множество ассоциативных отношений с другими концептами. С помощью алгоритма CONCOR<sup>9</sup> возможно увидеть, как в принципе осуществляется «дробление» текста на семантические группы. За счет того, что концепты в ORA-Lite можно двигать по виртуальной плоскости, они были аккуратно разнесены в разные стороны, чтобы показать, сколько слов мы можем ассоциировать с этими терминами, которые обозначают *одно и то же*. Примечательно, что с самого начала наблюдается пара: «*инфляция ускорилась*» и «*ускорение роста цен*». Однако более детализированный структурный анализ обращает внимание на то, что по ходу текста концепт *рост цен* (степень центральности 0,505; посредничества<sup>10</sup> — 0,171), скажем так, «работает» для отсылки к продуктам питания и товарам (которые более востребованы массовым читателем — например жителем страны), в то время как *инфляция* (степень центральности 0,479; посредничества — 0,091) «обслуживает» описание общей ситуации в экономике страны. По сути, метрики центральности, посредничества и общей центральности позволяют подтвердить на этапе контент-анализа, что ведущей темой текста является проблема роста цен, выраженная в виде двух концептов.

Перейдя непосредственно к схеме контент-анализа, можно обратить внимание на первый абзац новости: «*В сентябре инфляция ускорилась до 0,8% в сравнении с августом, годовая подскочила до 7%, сообщил Росстат. Ускорение роста цен перекинулось на непродовольственный сегмент, дорожавший вдвое быстрее, чем в среднем в предыдущие месяцы. Это приговор, считают эксперты: официальный прогноз — 8% за год — невыполним*». Этот отрывок текста хорошо отражает пример фрейминга события [11; 25], а также специфику драматизации сюжета [3]. Уже здесь можно было бы кодировать новые и очень любопытные категории и концепты: *инфляция, ускорилась, 0,8%, 7%, Росстат, ускорение, рост цен, перекинулось, вдвое быстрее, приговор, эксперты*.

<sup>9</sup> Мы получали схожие кластерные группы при использовании алгоритма Гирвана-Ньюмана [23]. Здесь и далее рекомендуется посмотреть цветные рисунки с А по F (для Текстов 1–6 соответственно) в электронном приложении к статье по адресу: URL: <https://www.jour.fnisc.ru/index.php/socjour/article/view/7262/7199>

<sup>10</sup> Здесь и далее более полное представление метрик для Текстов 1–6 дано в таблице в электронном приложении к статье по указанному адресу. Дискуссионным остается вопрос, насколько ряд метрик применим к тестовым вершинам. Существующие исследования позволяют это делать.

Конечно, это не означает, что для контент-анализа необходимо взять их все в кодировочную таблицу, однако такие концепты, как *Росстат*, *эксперты*, могут возникать не один раз. Остальные категории можно будет внедрять в логику анализа уже отдельно взятых текстов.

Безусловно, для полноты понимания текста необходимо добиться того, чтобы группы связанных между собой («близких») концептов и категорий были отчетливо видны. С точки зрения ССП желательнее выстраивать семантическую карту в виде кольцевых структур, где кластерная группировка упомянутым (к примеру) алгоритмом CONCOR дает лучшее представление о том, на какие микрофреймы делится текст. Например, не один раз появляется акцент на *сентябре* и *росте цен* (которые имеют множество «выходов» на другие концепты, иные смыслы). Интерпретации и контексты действительно связаны с концептом *рост цен*; более того, если исключить его из текста, то можно получить менее насыщенную смыслами интерпретацию (на что косвенно указывает степень брокерства концепта, равная 0,910 — чуть больше, чем у *инфляции* — 0,870). *Сентябрь* как концепт имеет множество ассоциаций с другими, в связи с чем является наиболее заметным с точки зрения степени центральности. Он располагается на «первом круге» и входит в одну группу семантической схожести с центральным концептом. На том же круге можно увидеть еще, к примеру, концепты *ускоряться*, *товар* (входит в другую группу) и т. п. Также у *роста цен* есть прямая отсылка к отдельным видам продуктов, что видно в оригинальном тексте: «К подорожавшему из-за летней жары и засухи зерну, спровоцировавшему рост цен на крупы, яйца, молочную продукцию, добавился картофель (+8,7% за месяц)». При этом *картофель* и *молочная продукция* находятся на «втором круге» (скорее всего, вследствие физического удаления от концепта *рост цен*), хотя в предложении они могут быть так же значимы, как и упоминание крупы. Что же касается *инфляции*, то у нее выделяются два момента: 1) принадлежность к отличной от *роста цен* семантической группе; 2) отсутствие прямой связи с *ростом цен*. С точки зрения обоснованной теории, мы должны были бы их кодировать как одно целое, но данное предположение на деле с помощью ССП не подтверждается: *инфляция* и *рост цен* имеют ассоциированные связи со многими одинаковыми концептами, но в одной группе не оказываются. При подстройке разными параметрами (например, уменьшением числа групп алгоритмом CONCOR с 6 до 2 [23]) мы получаем эти два концепта в одной группе.

Что касается остальных текстов, то можно обратить внимание на следующие моменты<sup>11</sup>. Текст 2 («Ведомости», 2014 г., написан тем же автором) значительно больше по объему. В нем в качестве центрального

---

<sup>11</sup> Как отмечалось выше, в Приложении приведена таблица с метриками для Текстов 4–6.

концепта выступает *инфляция*, в то время как *рост цен*, также не связанный с ней напрямую, находится в некоторой «родственной» зоне, что косвенно определяет количество ассоциаций с другими концептами. Это можно заметить по ряду сетевых метрик, где они находятся достаточно высоко и рядом друг с другом: например, центральность собственного вектора у концепта «инфляция» составляет 0,488, а у концепта «рост цен» 0,335). Второй интересный момент: в сетевой структуре было замечено, что концепты, отражающие числовые значения макростатистики, оказываются рядостоящими с концептом «инфляция», что косвенно может указывать на их смысловую взаимозависимость. Благодаря этому мы можем кодировать цифры вроде 0, 2015, 10 и др. как связующие звенья *инфляции*. В оригинальном тексте можно найти этому подтверждения: «*Инфляция обновила пятилетний рекорд: 10,5%. Стагфляция, обрушив уровень жизни населения, в 2015 г. может из экономической проблемы стать политической. <...> По расчетам Прилепского, к концу I квартала инфляция может достичь 16%, к концу 2015 г. вернется на уровень этого года — 10,5–10,8%*». А *рост цен* является локальным концептом, который связан со *стагфляцией*, *населением*, *экономическим кризисом*, *спадом* и напрямую — только с концептом *2015* («...в кризисном 2015 году *рост цен* может и не замедлиться...»).

Текст 3, сюжет 2010 года от Gazeta.ru, оказался по своей структуре схожим с Текстом 1, что говорит об общей устойчивой стратегии структурирования экономической информации, сопряженной с временным периодом. В этом можно убедиться, если рассмотреть рисунок 2 в Приложении. Однако, в отличие от кейса «Ведомостей», здесь мы фиксируем более выраженную смешанную стратегию, не только в плане установки повестки дня (для чего использование семантических карт оказывается очень важным). *Инфляция* и *рост цен* находятся практически в схожих положениях при расчете центральности посредничества, однако с позиции собственного вектора *рост цен* становится таким же значимым для текста в целом, как и *инфляция*, но на разных (по всей видимости) уровнях. У *инфляции* значимо больше связей с другими концептами (степень центральности 0,781) и она оказывается более значимой по содержанию (центральность собственного вектора 0,730). Удивительны два момента. Во-первых, сетевая структура оказалась такой, что *эксперты*, наиболее видимые при ручном кодировании текста, заняли место в сети «на втором плане». Во-вторых, *рост цен* с точки зрения кластерной структуры оказывается «скудным» на связи с другими концептами.

Текст 4 (Gazeta.ru) отличается тем, что в нем *инфляция* практически не встречается. Это позволяет предположить, что содержание текста изменилось с учетом интересов аудитории, которая может не воспринимать сложные экономические термины (в силу, к примеру, образования либо изменения повестки дня [3]).

Тексты 5 и 6 — публикации в «Комсомольской правде» (2010 и 2014 гг.) — пример стратегии структурирования экономической информации, где акцент сделан на самых простых концептах, а сложные максимально «отодвигаются». Этим они не похожи на предыдущие случаи, поскольку здесь предполагается максимальная адаптация текста для неподготовленного читателя. Семантическая структура подтверждает данный тезис: в центре внимания оказываются *магазины, продукты питания, цены, стоимость* и т. п. Отличительная черта Текста 5 — наличие большого количества разорванных цепочек слов. Автор новости в ряде случаев активно применял короткие предложения, что в итоге повлекло за собой «отрыв» части концептов от основной сети. К слову, его сюжет — это хороший пример интеграции *инфляции* в логику других концептов (*продукты питания, бензин, дизель* — находящиеся в разных группах). Метрика собственного вектора позволила продемонстрировать устойчивость выделенных главных концептов — ассоциациями в таком случае выступают *продукты, цены, топливо*. Числовых характеристик небольшое количество, и они буквально «раскиданы» по периферической зоне семантической сети.

Отличительная черта Текста 6 — акцент на образе курса рубля, где место *инфляции* и *роста цен* сугубо контекстуальное («*Импортёры вынуждены закладывать в свои цены риски дальнейшей девальвации и пытаются отыграть уже полученные потери, что подстегивает инфляцию. А последствия сильной инфляции, к сожалению, хорошо известны. Действительно, рост цен в стране резко ускорился. Как считают аналитики, ослабление валюты на 10% приводит к увеличению инфляции примерно на 1–1,5%. Так, по итогам этого года цены могут вырасти более чем на 9%, хотя изначальный прогноз был 5–6%*»).

### **Заключение**

В исследовании было рассмотрено применение семантического сетевого подхода для решения задач структурного анализа текстов, в частности экономической тематики. Основной единицей анализа служила «экономическая информация». Под этим можно понимать сложные термины, процессы, элементы массовой информации, во многом определяющей потребительские и финансовые решения простых людей. Поскольку не все люди располагают профильными знаниями по экономике, агенты СМИ могут предлагать разные способы реконструкции сюжета экономических событий, адаптируя при этом как текст, так и ассоциации, необходимые для «припоминания» читателем недавних событий и обстоятельств. В рассмотренных кейс-стади встречались два текста одного автора, написанные с интервалом в четыре года. И спустя четыре года была замечена небольшая смена стратегии, от более общего, «объективного» видения (с акцентом на «инфляции») к контекстуальной паре «инфляция — рост цен». Это может положительно влиять на качество усвоения новостного материала.

В текстах Gazeta.ru мы увидели разные подходы к наполнению сюжета содержанием. Однако в двух разных текстах видна общая стратегия использования «смешанной» тактики применения слова «инфляция», тогда как чаще читателю преподносятся термин «рост цен» и создается описание сюжета уже вокруг него. При этом Текст № 3 был специфичен потому, что изначально речь идет про *глобальный рост цен*, хотя далее по тексту фокус смещается на определенную страну, а сюжеты «Комсомольской правды» оказались более «жизненными», поскольку автор стремился «простыми» концептами описать примерно то же, что и автор первых двух кейсов. В рамках анализа было продемонстрировано, что новость представляется как минимум в двух структурах: экономическая информация описывается с «официальной» стороны и с «жизненной». В таком случае ощущается стремление представить инфляцию как самостоятельный термин, со сложным аналитическим содержанием и интерпретацией; однако это еще и попытка прояснить ее роль через прогнозы, ассоциации и апелляцию к экспертной точке зрения. Специализированные издания будут освещать события вокруг инфляции с официальной позиции, в то время как массовые — с «жизненной» стороны.

Что нам дает работа с методологией ССП? Во-первых, с помощью разных способов визуализации мы можем по-разному увидеть взаимное расположение концептов в сетевой структуре, буквально «вращая» карту в разных плоскостях. Разная картинка демонстрирует различное понимание влияния отдельных концептов на смысловой дискурс всего текста, позволяя расставить акценты и приоритеты для читателя. Понятно, что «Шок на год» — обзорная статья, в которой необходимо учитывать сугубо макроэкономические явления и интерпретацию в терминах действий (например «рост цен»). А дополняющие картину количественные данные (см. табл. 3) позволяют увидеть, как производится кодирование текстов. Похожую картину можно наблюдать и в других сообщениях. В этом плане ССП выступает для исследователя в роли *пред-критерия* при осуществлении контент-анализа.

Во-вторых, сам по себе сетевой анализ работает с математическим аппаратом, основанным на принципах дискретной математики, а объектом выступает единица, включенная в социальные процессы. Результат — это представление социальной структуры в виде графов — ее визуализация. Оценка графов основывается на метриках, характеризующих положение каждого концепта в пространстве, общую устойчивость всего графа и т. п. В пространстве компьютерных алгоритмов (LDA) ССП может выступить в качестве *пост-критерия* оценки точности выделенных тем, поскольку у исследователя может возникнуть необходимость в выработке объяснительной модели того, как «машина» выделяет основные концепты, ключевые слова в качестве ведущих.

Главный тезис сказанного выше в том, что сетевой анализ — «гибридный» метод анализа текстовой информации. Существуют исследования, которые хорошо демонстрируют активное применение

именно ССП как ведущего метода<sup>12</sup>. В социологических исследованиях можно не только изучать структурные позиции индивидов в одной совокупности, но и определять властные отношения, плотность всей сети взаимодействий и многое другое.

Однако отметим несколько существенных ограничений, которые были выявлены в ходе проведения небольшого эксперимента. Так, исследователю нужно заранее определять фокус анализа: либо делать акцент на каждом новостном тексте отдельно, либо проводить кросс-текстовый анализ путем совмещения всех собранных текстов в одну крупную семантическую сеть, а это потенциально может привести к смешению сразу нескольких подходов [18; 20]. Если использовать ССП совместно с другими методами, особенно с методологией контент-анализа, то рассматриваемый подход необходимо специфицировать по его роли в зависимости от типа исследования: если речь о количественном контент-анализе, то, вероятно, его имеет смысл использовать после процедуры метода, чтобы иметь «подтверждающее» представление о выделенных концептах. Особый вопрос возникает в части дискурс-анализа, где потребуется приобрести дополнительные навыки компьютерной обработки текстов, а также совмещения с инструментами контент-анализа (Atlas.Ti). Количественные данные можно внедрять в регрессии, но спецификации модели в таком случае потребуется уделять особое внимание. Вторичным ограничением ССП является привязка к ряду существующего программного обеспечения.

#### ЛИТЕРАТУРА

1. Батура Т.В. Семантический анализ и способы представления смысла текста в компьютерной лингвистике // Программные продукты и системы. 2016. № 4 (116). С. 45–57. DOI: 10.15827/0236-235X.114.045-057
2. Двора Я., Ван Хульст М., Ткачева К., Вахштайн В. Фреймы политического: от фрейм-анализа к анализу фреймирования // Социологическое обозрение. 2011. Т. 10. № 1–2. С. 87–113.
3. Казаков А.А. Теория установления повестки дня vs фрейминг: к вопросу о соотношении подходов // Полития. 2015. № 1 (76). С. 103–113.
4. Контент-анализ СМИ: проблемы и опыт применения / Под ред. В.А. Мансурова. М.: Институт социологии РАН, 2010. — 324 с., 1 CD-ROM.
5. Кориунов А., Гомзин А. Тематическое моделирование текстов на естественном языке // Труды института системного программирования РАН. 2012. Т. 23. С. 215–244.
6. Лукьянова Г.В. Дискурс-анализ текстов СМИ // Политический анализ. 2009. № 9. С. 154–165.

---

<sup>12</sup> См., например: MyStem: инструмент для морфологического анализа текста // Яндекс [электронный ресурс]. Дата обращения: 02.04.2020. URL: <https://yandex.ru/dev/mystem/>

7. Тертычный А.А. Характер применения жанров в современных печатных СМИ // Вопросы теории и практики журналистики. 2012. № 2. С. 106–112.
8. Холмогорова В. Россияне назвали бедность, цены и безработицу главными проблемами // РБК. 31 августа 2017 [электронный ресурс]. Дата обращения 01.12.2018. URL: <https://www.rbc.ru/politics/31/08/2017/59a68cdf9a79477acee3d38f>
9. Экономика и ее освещение в СМИ: об экономической ситуации, ее динамике и освещении положения дел в экономике в СМИ // ФОМ. 30 августа 2018 [электронный ресурс]. Дата обращения 01.06.2019. URL: <https://fom.ru/Ekonomika/14093>
10. Brachman R.J. What's in a concept: Structural foundations for semantic networks // International journal of man-machine studies. 1977. Vol. 9. No. 2. P. 127–152. DOI: 10.1016/s0020-7373(77)80017-5
11. Cacciatore M.A., Scheufele D.A., Iyengar S. The end of framing as we know it... and the future of media effects // Mass Communication and Society. 2016. Vol. 19. No. 1. P. 7–23. DOI: 10.1080/15205436.2015.1068811
12. Carley K. Coding choices for textual analysis: A comparison of content analysis and map analysis // Sociological methodology. 1993. P. 75–126. DOI: 10.2307/271007
13. DiMaggio P., Nag M., Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding // Poetics. 2013. Vol. 41. No. 6. P. 570–606. DOI: 10.1016/j.poetic.2013.08.004
14. Drieger P. Semantic network analysis as a method for visual text analytics // Procedia-social and behavioral sciences. 2013. Vol. 79. P. 4–17. DOI: 10.1016/j.sbspro.2013.05.053
15. Eddington S.M. The communicative constitution of hate organizations online: A semantic network analysis of “Make America great again” // Social Media + Society. 2018. Vol. 4. No. 3. P. 1–12. DOI:10.1177/2056305118790763
16. Evans J.A., Aceves P. Machine translation: Mining text for social theory // Annual Review of Sociology. 2016. Vol. 42. P. 21–50. DOI: 10.1146/annurev-soc-081715-074206
17. Gamson W.A., Croteau D., Hoynes W., Sasson T. Media images and the social construction of reality // Annual review of sociology. 1992. Vol. 18. No. 1. P. 373–393. DOI: 10.1146/annurev.so.18.080192.002105
18. Grimmer J., Stewart B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts // Political analysis. 2013. Vol. 21. No. 3. P. 267–297. DOI: 10.1093/pan/mps028
19. Guo L. The application of social network analysis in agenda setting research: A methodological exploration // Journal of Broadcasting & Electronic Media. 2012. Vol. 56. No. 4. P. 616–631. DOI: 10.1080/08838151.2012.732148
20. Jacobi C., Van Atteveldt W., Welbers K. Quantitative analysis of large amounts of journalistic texts using topic modelling // Digital Journalism. 2016. Vol. 4. No. 1. P. 89–106. DOI: 10.1080/21670811.2015.1093271

21. *Krippendorff K.* Content Analysis: An Introduction to its Methodology. Thousand Oaks, CA: Sage. 2004. — 413 p.
22. *Lehman F.* Semantic networks // Computers and Mathematics with Applications. 1992. Vol. 23. No. 2–5. P. 1–50.
23. *Newman M., Barabasi A.L., Watts D.J.* The structure and dynamics of networks. Princeton University Press, 2006. — 592 p.
24. *Schultz F., Kleinnijenhuis J., Oegema D., Utz S., Van Atteveldt W.* Strategic framing in the BP crisis: A semantic network analysis of associative frames // Public Relations Review. 2012. Vol. 38. No. 1. P. 97–107. DOI: 10.1016/j.pubrev.2011.08.003
25. *Short J.C., McKenny A.F., Reid S.W.* More than words? Computer-aided text analysis in organizational behavior and psychology research // Annual Review of Organizational Psychology and Organizational Behavior. 2018. Vol. 5. P. 415–435. DOI: 10.1146/annurev-orgpsych-032117-104622
26. *Simoff S., Böhlen M.H., Mazeika A.* (eds.) Visual data mining: Theory, techniques and tools for visual analytics. Springer Science & Business Media, 2008. Vol. 4404. — X, 407 p. DOI: 10.1007/978-3-540-71080-6
27. *Sowa J.F.* (ed.). Principles of semantic networks: Explorations in the representation of knowledge. Burlington: Morgan Kaufmann, 2014. — 594 p.
28. *Tichenor P.J., Donohue G.A., Olien C.N.* Mass media flow and differential growth in knowledge // Public opinion quarterly. 1970. Vol. 34. No. 2. P. 159–170. DOI: 10.1086/267786
29. *Valkenburg P.M., Peter J., Walther J.B.* Media effects: Theory and research // Annual review of psychology. 2016. Vol. 67. P. 315–338.
30. *Van Dijk T.A.* Critical discourse studies: A sociocognitive approach // Methods of critical discourse analysis. 2009. Vol. 2. No. 1. P. 62–86.

## ПРИЛОЖЕНИЕ

## Таблица

**Количественная характеристика работы ведущих концептов в текстах**

	Текст № 4	Текст № 5	Текст № 6
Число концептов (фактическое)	393	367	398
Число связей (фактическое)	1062	1003	985
Плотность сети, %	1,38%	1,49%	1,25%
Доля успешной кластеризации, %	50,60%	54,60%	56,90%
Число устойчивых смысловых групп	535	509	477

	Текст № 4	Текст № 5	Текст № 6			
Степень центральности (10 концептов с самым большим числом связей в доле от всего текста)	цена	0,836	год	0,807	рубль	0,809
	повышение	0,549	цена	0,794	курс	0,635
	фас	0,458	рубль	0,457	валюта	0,502
	продукт	0,441	вырастать	0,379	год	0,484
	стоимость	0,418	акциз	0,323	доллар	0,415
	рост_цен	0,323	водка	0,316	говорить	0,338
	продукция	0,269	товар	0,307	ослабление	0,293
	хлеб	0,242	продукт	0,295	рынок	0,278
	мясо	0,226	сейчас	0,294	ЦБ	0,272
	декабрь	0,222	становиться	0,281	компания	0,248
Степень собственного вектора (наибольшая выраженная важность 10 концептов в сети)	цена	0,742	цена	0,603	рубль	0,733
	повышение	0,655	год	0,501	курс	0,629
	продукт	0,351	вырастать	0,342	говорить	0,353
	стоимость	0,32	водка	0,217	доллар	0,267
	хлеб	0,237	сейчас	0,204	ослабление	0,238
	необоснованный	0,229	средний	0,203	валюта	0,211
	мясо	0,218	рубль	0,202	ЦБ	0,164
	возможный	0,192	товар	0,2	рынок	0,154
	например	0,172	зима	0,194	приводить	0,146
	гречка	0,162	назад	0,188	считать	0,129

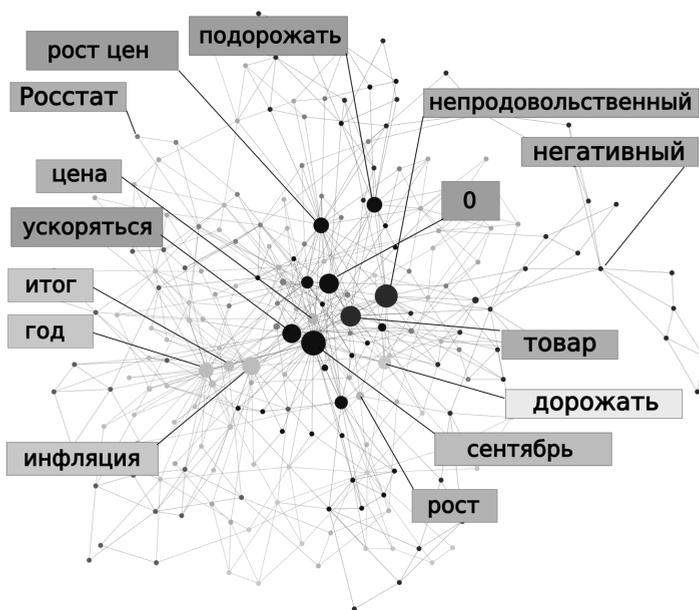
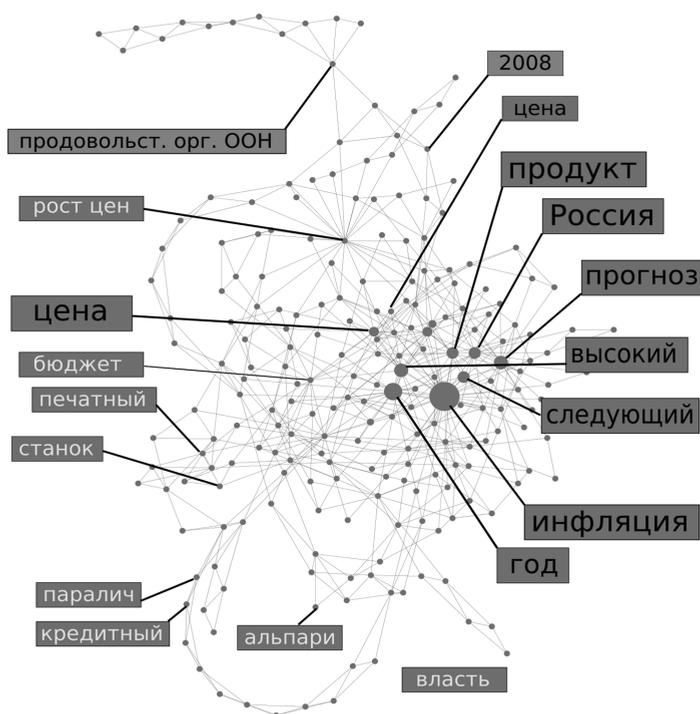


Рис. 1. Графическое представление семантической структуры для Текста № 1



**Рис. 2. Графическое представление семантической структуры для Текста № 3**

Дата поступления: 11.02.2019.

---

**SOTSIOLOGICHESKIY ZHURNAL = SOCIOLOGICAL JOURNAL. 2020.**  
**VOL. 26. No. 2. P. 8–30. DOI: 10.19181/socjour.2020.26.2.7262**

**S.G. PASHKOV**

National Research University “Higher School of Economics”,  
 Moscow, Russian Federation.

**Stanislav G. Pashkov** — Postgraduate Student, Lecturer, Faculty of Social Sciences;  
 Research Assistant, Laboratory for Studies in Economic Sociology, National Research  
 University Higher School of Economics. **Address:** 11, Myasnitskaya str., 101000, Moscow,  
 Russian Federation. **Phone:** +7 (495) 772-95-90\*12450. **Email:** spashkov@hse.ru

**THE SEMANTIC NETWORK APPROACH: OPPORTUNITIES AND RESTRICTIONS  
 (EXAMPLE OF INFLATION IMAGE IN THE MEDIA)**

*Abstract.* This article focuses on there being a need for tools that can facilitate coding and analysis processes for news reports. The study was based on a set of economic news replete with specific terms, interpretations, expertise and metaphorical description of events. In most

cases it can be argued that the content of the texts is complicated, thus “classical” content analysis may require additional iterations and increased attention to the analytical procedure. The study highlights the methodological, analytical features of the semantic network approach (SNA) in comparison with the content analysis and Text Mining approaches based on analyzing six economic news items containing the terms “*rising prices*” and “*inflation*”. SNA is distinguished by simplification of large unstructured data processing with emphasis on content. The preparation and calculation of network metrics for each news report leads to the most significant concepts being reflected. That simplifies the content analysis of a larger body of texts. In several cases visualization shows different semantic positions of “inflation” being a synonym for “rising prices” depending on the topic. As an important result, regardless of the volume and visual structure of the news message, these terms can be considered as leading in the corresponding storylines that can help conduct a discourse analysis with their mention. It is assumed that this approach will become a “supporting” tool for further quantitative and qualitative analysis of news reports, particularly on economic topics. The technical features of text preparation and semantic modeling programs can be considered as potential limitations of the approach, especially in the space of Text Mining.

*Keywords:* socio-economic information; news texts; semantic network approach; semantic maps; text processing; content analysis; discourse.

**For citation:** Pashkov S.G. The Semantic Network Approach: Opportunities and Restrictions (Example of Inflation Image in the Media). *Sotsiologicheskii Zhurnal = Sociological Journal*. 2020. Vol. 26. No. 2. P. 8–30. DOI: 10.19181/socjour.2020.26.2.7262

## REFERENCES

1. Batura T.V. Semantic analysis and methods of text meaning representation in computer linguistics. *Programmnye produkty i sistemy*. 2016. No. 4 (116). P. 45–57. (In Russ.) DOI: 10.15827/0236-235X.114.045-057
2. Dvora Ya., Van Khul'st M., Tkacheva K., Vakhshstain V. Dvora Ya., Van Hulst M. The Political/Process Promise of Policy Framing. *Sotsiologicheskoe obozrenie*. 2011. Vol. 10. No. 1–2. P. 87–113. (In Russ.)
3. Kazakov A.A. The Agenda-Setting Theory vs framing: On the question of the relationship of approaches. *Politiya*. 2015. No. 1 (76). P. 103–113. (In Russ.)
4. *Kontent-analiz SMI: problemy i opyt primeneniya*. [Content Analysis of Media: Problems and Application Experience.] Ed. by V.A. Mansurov. Moscow: Institut sotsiologii RAN publ., 2010. 324 p., 1 CD-ROM. (In Russ.)
5. Korshunov A., Gomzin A. Thematic modeling of texts in natural language. *Trudy instituta sistemnogo programmirovaniya RAN*. 2012. Vol. 23. P. 215–244. (In Russ.)
6. Luk'yanova G.V. Discourse Analysis of Media Texts. *Politicheskii analiz*. 2009. No. 9. P. 154–165. (In Russ.)
7. Tertychnyi A.A. The nature of the use of genres in Modern Print Media. *Voprosy teorii i praktiki zhurnalistiki*. 2012. No. 2. P. 106–112. (In Russ.)
8. Kholmogorova V. Russians called poverty, prices and unemployment the main problems. *RBC*. 31.08.2017. Accessed 01.12.2018. URL: <https://www.rbc.ru/politics/31/08/2017/59a68cdf9a79477acee3d38f> (In Russ.)
9. Economics and its media coverage: on the economic situation, its dynamics and coverage of the state of the economy in the media. *FOM*. 30.08.2018. Accessed 01.06.2019. URL: <https://fom.ru/Ekonomika/14093> (In Russ.)
10. Brachman R.J. What's in a concept: Structural foundations for semantic networks. *International journal of man-machine studies*. 1977. Vol. 9. No. 2. P. 127–152. DOI: 10.1016/s0020-7373(77)80017-5

11. Cacciatore M.A., Scheufele D.A., Iyengar S. The end of framing as we know it... and the future of media effects. *Mass Communication and Society*. 2016. Vol. 19. No. 1. P. 7–23. DOI: 10.1080/15205436.2015.1068811
12. Carley K. Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological methodology*. 1993. P. 75–126. DOI: 10.2307/271007
13. DiMaggio P., Nag M., Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*. 2013. Vol. 41. No. 6. P. 570–606. DOI: 10.1016/j.poetic.2013.08.004
14. Drieger P. Semantic network analysis as a method for visual text analytics. *Procedia-social and behavioral sciences*. 2013. Vol. 79. P. 4–17. DOI: 10.1016/j.sbspro.2013.05.053
15. Eddington S.M. The communicative constitution of hate organizations online: A semantic network analysis of “Make America great again”. *Social Media + Society*. 2018. Vol. 4. No. 3. P. 1–12. DOI:10.1177/2056305118790763
16. Evans J.A., Aceves P. Machine translation: Mining text for social theory. *Annual Review of Sociology*. 2016. Vol. 42. P. 21–50. DOI: 10.1146/annurev-soc-081715-074206
17. Gamson W.A., Croteau D., Hoynes W., Sasson T. Media images and the social construction of reality. *Annual review of sociology*. 1992. Vol. 18. No. 1. P. 373–393. DOI: 10.1146/annurev.so.18.080192.002105
18. Grimmer J., Stewart B.M. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*. 2013. Vol. 21. No. 3. P. 267–297. DOI: 10.1093/pan/mps028
19. Guo L. The application of social network analysis in agenda setting research: A methodological exploration. *Journal of Broadcasting & Electronic Media*. 2012. Vol. 56. No. 4. P. 616–631. DOI: 10.1080/08838151.2012.732148
20. Jacobi C., Van Atteveldt W., Welbers K. Quantitative analysis of large amounts of journalistic texts using topic modeling. *Digital Journalism*. 2016. Vol. 4. No. 1. P. 89–106. DOI: 10.1080/21670811.2015.1093271
21. Krippendorff K. *Content Analysis: An Introduction to its Methodology*. Thousand Oaks, CA: Sage, 2004. 413 p.
22. Lehman F. Semantic networks. *Computers and Mathematics with Applications*. 1992. Vol. 23. No. 2–5. P. 1–50.
23. Newman M., Barabasi A.L., Watts D.J. *The structure and dynamics of networks*. Princeton University Press, 2006. 592 p.
24. Schultz F., Kleinnijenhuis J., Oegema D., Utz S., Van Atteveldt W. Strategic framing in the BP crisis: A semantic network analysis of associative frames. *Public Relations Review*. 2012. Vol. 38. No. 1. P. 97–107. DOI: 10.1016/j.pubrev.2011.08.003
25. Short J.C., McKenny A.F., Reid S.W. More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*. 2018. Vol. 5. P. 415–435. DOI: 10.1146/annurev-orgpsych-032117-104622
26. Simoff S., Böhlen M.H., Mazeika A. (eds.) *Visual data mining: Theory, techniques and tools for visual analytics*. Springer Science & Business Media, 2008. Vol. 4404. X, 407 p. DOI: 10.1007/978-3-540-71080-6
27. Sowa J.F. (ed.). *Principles of semantic networks: Explorations in the representation of knowledge*. Burlington: Morgan Kaufmann, 2014. 594 p.
28. Tichenor P.J., Donohue G.A., Olien C.N. Mass media flow and differential growth in knowledge. *Public opinion quarterly*. 1970. Vol. 34. No. 2. P. 159–170. DOI: 10.1086/267786
29. Valkenburg P.M., Peter J., Walther J.B. Media effects: Theory and research. *Annual review of psychology*. 2016. Vol. 67. P. 315–338.
30. Van Dijk T.A. Critical discourse studies: A sociocognitive approach. *Methods of critical discourse analysis*. 2009. Vol. 2. No. 1. P. 62–86.