

## Прогнозирование научной области (на материале ведущего тематического журнала)\*

*Описывается исследовательская программа моделирования, оценки и прогнозирования состояния информационного пространства, сложившегося в отдельной предметной области, которая формируется ведущим научным журналом. В качестве предметной области рассмотрена когнитивная лингвистика, а материалом исследования послужили публикации российского научного журнала “Вопросы когнитивной лингвистики”, целиком посвященного разработке лингвокогнитивной проблематики.*

**Ключевые слова:** научная предметная область, научный журнал, научная статья, термин, терминополь, графосемантическое моделирование, статистические методы, информационная система “Семозраф”

### ВВЕДЕНИЕ

Научная предметная область (далее – ПрО) представляет собой создаваемую индивидуальными и коллективными агентами научного производства открытую мультиструктурную информационную систему, в которой осуществляется непрерывный интра- и интерпредметный процесс обмена информацией по доступным информационным каналам. Результатом информационных процессов становится постоянное обновление содержательной и структурной составляющих научной ПрО: конкуренция (изменение конфигураций, зависимостей, влияния друг на друга) частнонаучных ПрО в рамках общей предметной области, а также обновление концептуального пространства ПрО за счет междисциплинарных связей.

Научный журнал – одна из ключевых подсистем ПрО, являющаяся основным каналом обмена научной информацией, платформой для структурирования самой ПрО, а также средством

оценки эффективности научной деятельности и ценности ее результатов по отношению к агентам научного производства [1, 2]. Научные журналы становятся составной частью технологии выделения финансовой поддержки со стороны государства и частных научных фондов.

Несмотря на интенсификацию исследований в области моделирования ПрО науки (экономики, медицины и др.), научные журналы интересуют исследователей, как правило, в аспекте публикационной активности, а не того информационного поля, кото-

рое они формируют. Основу расчетов разнообразных индексов (цитируемости, самоцитируемости, Хирша, импакт-фактора и мн. др.) составляют сугубо формальные количественные показатели числа публикаций и цитирований за определенные периоды времени. Эти формальные показатели используются как основа планирования научной деятельности, в том числе в аналитических инструментах InCites и SciValSpotlight (приложениях, работающих на базе наукометрических показателей WoS и Scopus), созданных для оценки текущей деятельности агентов научного производства и применяемых как инструмент принятия решения в области финансирования проектов и научных коллективов [3–5].

Формальные инструменты дают показатели эффективности *результатов* научной деятельности, но не могут быть положены в основу планирования исследовательской деятельности в прогнозируемом состоянии ПрО, так как они не связаны с собственно информационным пространством ПрО. Этим обусловлена актуальность поисков новых форм, методов, инструментов моделирования информационного пространства, сложившегося в отдельной научной ПрО, для эффективного планирования научной деятельности индивидуальных и коллективных агентов научного производства. В этой связи представляют интерес модели информационного пространства ведущих научных журналов в исследуемой ПрО, так как ведущие научные журналы являются: а) главным “инструментом” разработки ПрО со стороны научного сообщества, б) средством ориентации исследователей, занимающихся разработкой отдельных аспектов ПрО, в самых последних разработках; в) средством диагностики и мониторинга состояния ПрО, г) конфигуратором частнонаучных ПрО (опре-

\* Исследование выполнялось при финансовой поддержке РГНФ (проект № 13-34-01276) и РФФИ (проект РФФИ № 14-06-31143).

деляющим “вес” и положение отдельных направлений, школ, концепций относительно друг друга в рамках общей ПрО), д) проводником новых идей и направлений. Научный тематический журнал имеет не только синхронное измерение, но и диахронное, он представляет историю развития ПрО. Результаты исследования организации информационного пространства научного журнала (в том числе и в сопоставлении с организацией информационного пространства других научных журналов, в первую очередь зарубежных) могут быть полезны самым разным группам агентов научного производства.

## МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Предлагаемая исследовательская программа изучения частнонаучных ПрО основывается на моделировании терминосферы ПрО. В качестве материала выступает корпус научных текстов, относящихся к ПрО.

В настоящей статье в качестве ПрО рассмотрена КОГНИТИВНАЯ ЛИНГВИСТИКА, а непосредственным материалом послужили публикации авторитетного российского научного журнала “Вопросы когнитивной лингвистики”, целиком посвященного разработке лингвокогнитивной проблематики, с № 1 за 2008 г. по № 4 за 2013 г. (24 номера за 6 лет).

*Операциональными единицами* послужили базовые термины (в широком смысле слова), выносимые авторами публикаций в наборы ключевых слов (далее – НКС) к своим статьям. Обращение к ключевым словам (далее – КС) научных публикаций обусловлено тем, что, во-первых, посредством КС и НКС авторы сами обозначают доминанты концептуального пространства своих исследований; во-вторых, НКС научных публикаций представляют собой легко формализуемый конструкт в рамках большого корпуса текстов; в-третьих, НКС, в отличие от статьи, к которой они относятся, обычно доступны для автоматизированного извлечения.

В то же время необходимо остановиться на проблеме статуса КС и перспектив их использования в качестве операциональных единиц в нашем исследовании.

КС в научной статье представляют собой отдельные лексемы или словосочетания, *значительную часть которых нельзя отнести ни к терминосистеме, ни даже к терминологии изучаемой ПрО.* Одна (причем довольно объемная) часть КС к научным статьям представляет собой общенаучный слой терминологии (КАТЕГОРИЯ, СИСТЕМА, ПОНЯТИЕ и др.), другая содержит специальную лексику, характерную либо для языкознания в целом (ЯЗЫК, СЛОВО, ПРЕДЛОЖЕНИЕ и др.), либо для других его разделов/направлений (ДИСКУРС, ПОЛИТИЧЕСКАЯ КОММУНИКАЦИЯ, ИНТЕРТЕКСТУАЛЬНОСТЬ и др.). И, наконец, внушительная по объему часть КС относится не к терминологии, а к номенклатуре (ЛЕЗГИНСКИЙ ЯЗЫК, ЛЕКСЕМА “НОЧЬ”, ОБРАЗ ЛОШАДИ и др.).

Таким образом, НКС к научным публикациям представляют собой предметно-понятийный субстрат (термины, номены, онимы), относящийся к разным подсистемам и структурам информационного про-

странства науки: терминологии и номенклатуре; общенаучным, дисциплинарным, частнонаучным (в том числе и смежным) ПрО. В НКС преломляются качества открытости, системности, полиструктурности, идущие одновременно диффузные и интеграционные процессы в самой науке.

Чтобы оценить соотношение терминологии ПрО (ее концептуального ядра) и сопутствующего номенклатурно-терминологического материала, достаточно сопоставить наши данные с известными корпусными исследованиями. Так, например, база данных ИНИОН по языкознанию [6], созданная на основе индексации 293440 документов, содержит в созданном на ее основе тезаурусе 1890 терминов (см. также [7]). В то же время материал нашего исследования включает 1650 КС к 446 научным публикациям.

Научная статья, репрезентирующая целостный фрагмент проведенного исследования, в полной мере обладает теми же качествами, что и “взраставшая” ее научная среда. Качества открытости, системности, диффузности и интегративности, полиструктурности проявляются в конструировании информационного пространства научной статьи и ее смысловых доминант, выносимых в композиционную область КС. Таким образом, термины, номены, онимы, не относящиеся напрямую к исследуемой ПрО, тем не менее крайне значимы для нее: *без этого предметно-понятийного субстрата невозможно представить ни синхронный срез ПрО, ни предложить ее ретроспективную или перспективную модель развития.*

Моделирование информационного пространства ПрО на основе множества понятий и реалий, репрезентированных в НКС, не в полной мере соответствует методам, используемым как в терминоведении, так и в онтологическом инжиниринге. В первую очередь речь идет о таком базовом методе исследования информационного пространства научной ПрО, как метод построения тезаурусов [8–11]. Тезаурус представляет собой модель языка науки, в которой связь между понятиями осуществляется “горизонтальными” и “вертикальными” семантическими типами отношений. В нашем случае сложность построения тезауруса обусловлена: 1) разнородностью материала, представляющего самые разнообразные стороны бытия науки, а не ПрО, которую он должен репрезентировать и структурировать; 2) неоднозначностью классификации – отнесения понятия к определенному узлу тезауруса. Сказанное в большой степени относится к терминологии и номенклатуре, так как термины и номены часто состоят из двух и более лексем, каждая из которых может быть отнесена к разным узлам тезауруса. Например, термин КОНЦЕПТУАЛЬНАЯ СТРУКТУРА ТЕКСТА был отнесен к полям КОНЦЕПТ И КОНЦЕПТОСФЕРА, (*концептуальная структура текста*), ТЕКСТ И ДИСКУРС (*концептуальная структура текста*) и СТРУКТУРНАЯ ОРГАНИЗАЦИЯ (*концептуальная структура текста*).

В то же время для нашего материала релевантен метод полевого анализа [12–14], с помощью которого структурируются семантические, понятийные, терминологические, тематические, ассоциативные и др. поля.

Остановимся на *принципах классификации КС*.

1. При отнесении термина (в широком смысле слова) к тому или иному терминополью рассматривалось наличие эксплицитной выраженности в семантике термина компонентов, связанных с семантикой полей. Например, в ключевом слове КОНЦЕПТ СВОЙ–ЧУЖОЙ эксплицитно представлен компонент КОНЦЕПТ, непосредственно входящий в терминополь КОНЦЕПТ И КОНЦЕПТОСФЕРА.

2. В случае отсутствия в семантике термина эксплицитных семантических компонентов, непосредственно входящих в те или иные поля, отнесение термина к полю осуществлялось с опорой на научную традицию, сложившуюся в лингвистике и смежных с ней науках. Например, ЦЕЛОСТНОСТЬ ВОСПРИЯТИЯ ТЕКСТА относится к терминополью РЕЧЕВАЯ ДЕЯТЕЛЬНОСТЬ на основании того, что проблемы восприятия речи обычно рассматриваются в рамках теории речевой деятельности.

3. В случае неоднозначности трактовки КС привлекается контекст его употребления: НКС, название статьи и аннотация.

4. Классификация осуществляется несколькими экспертами; вырабатывается согласованная позиция по спорным вопросам.

5. Одно понятие, выраженное посредством КС, может быть отнесено к одному или к нескольким полям одновременно (см. пример выше).

6. В качестве основного критерия ограничения содержания поля была взята частотность употребления в выборке текстов терминов, входящих в данное поле. В том случае, если частотность терминов, входящих в поле, достигала определенного уровня (определяемого статистически), данное терминополь обретало самостоятельность.

7. При формировании полей авторы стремились: а) по возможности сохранять определенность предметов исследований (т.е. не редуцировать поля до общих обозначений лингвистических направлений и парадигм), б) избегать излишней дробности предметов исследований (для сохранения целостного образа информационного поля).

8. Реализация пп. 6 и 7 приводит к тому, что в представленной ниже графосемантической модели ее отдельные терминополья находятся на разных уровнях иерархии ПрО. Например, поле ПСИХОЛОГИЧЕСКИЕ КАТЕГОРИИ связано с самыми разнообразными сторонами психологической жизни человека, изучаемыми психологической наукой. В то же время поля ЛЕКСИКА И ФРАЗЕОЛОГИЯ, ГРАММАТИКА и др. характеризуют либо уровни языковой системы/разделы лингвистики, либо языковые единицы/лингвистические объекты (ТЕКСТ И ДИСКУРС).

Однако, если с позиции формального подхода “неравное” положение терминопольей можно рассматривать как недостаток классификационной модели, то, на наш взгляд, оно скорее является достоинством, так как создаваемая модель стремится передать реальное положение вещей, эксплицируя “веса” научных проблем, актуальных в настоящее время в когнитивной лингвистике.

## СБОР И ПЕРВИЧНЫЙ АНАЛИЗ МАТЕРИАЛА

Исследование осуществлялось с опорой на *метод графосемантического моделирования*, реализованный в Информационной системе (далее – ИС) “Семограф” (<http://semograph.com>). Для построения графов использовалось программное средство Gephi (<http://gephi.org>). В качестве основного инструмента исследования графосемантической модели использовался R – язык программирования высокого уровня, предназначенный для статистической обработки данных (<http://www.r-project.org>).

Работа в ИС “Семограф” включает в себя создание Проекта (рабочего пространства, в котором осуществляется исследовательский цикл, реализованный в ИС “Семограф”). Фрейм Проекта состоит из следующих элементов:

1. Контекст – в нашем случае описание научной статьи. Контекст включает в себя аннотацию, название статьи и НКС.

2. Корпус контекстов (описание всех публикаций журнала “Вопросы когнитивной лингвистики”).

3. Набор значений, описывающих контекст (научную статью), в том числе:

3.1. Метаданные (тип текстовых данных, позволяющих ввести дополнительную информацию о статье: ФИО автора/авторов статьи; название статьи; научный статус автора/авторов статьи (аспирант, без ученой степени, кандидат наук, доктор наук); год публикации (2008, 2009, 2010, 2011, 2012; 2013); номер журнала (№1, №2, №3, №4); город, в котором проживает автор (спектр наименований).

Особенностью использования метаданных как типа переменных в ИС “Семограф” состоит в возможности множественной фильтрации по значениям метаданных и составления из всего собранного научного контента разнообразных выборок. Например, мы можем составить выборки контекстов научных статей, написанных только аспирантами или только докторами наук из Волгограда, опубликовавшими свои работы в 2009 г., в НКС которых используется термин “концепт” и/или “фрейм” и т.п. ИС “Семограф” предоставляет возможность исследовать семантическое пространство как всего корпуса контекстов, так и отдельной выборки.

3.2. Семантические компоненты (НКС к каждой статье). Например, статью “Фольклорный дискурс: когнитивно-дискурсивное исследование” [15, с. 50] характеризует следующий НКС: **ФОЛЬКЛОРНЫЙ ДИСКУРС, КОГНИТИВНО-ДИСКУРСИВНОЕ ИССЛЕДОВАНИЕ, КОММУНИКАТИВНЫЕ СТРАТЕГИИ, ЦЕННОСТИ, ДИСКУРСООБРАЗУЮЩИЕ КОНЦЕПТЫ.**

4. Семантические поля – в нашем случае терминополья, объединяющие отдельные научные понятия, выраженные с помощью КС. Терминополье репрезентирует входящие в него термины, что позволяет решить проблему вариативности КС (в том числе и терминологических повторов разного рода, например, синонимических). Для выполнения данного этапа требуется привлечение экспертов моделируемой научной отрасли. В задачу экспертов входит группирование доступных терминов (КС к статьям) в терминополья, формулирование названий терминопольей.

Отметим, что связи между полями и компонентами относятся к типу “многие ко многим”, т.е. возможно вхождение одного термина в несколько терминопольей и нескольких терминов в одно терминополье. Классификация терминов должна выполняться несколькими экспертами, т.к. существует вероятность субъективного отнесения терминов к терминопольям.

5. Графосемантическая модель, объединяющая вышеописанные элементы фрейма Проекта, может дополняться новыми контекстами и допускает изменение семантических полей и их связей с компонентами. На основе графосемантической модели строятся семантическая карта (С-карта) и семантический граф (С-граф).

5.1. С-карта отражает совместное присутствие двух терминопольей в одном и том же контексте с учетом подобной встречаемости во всех контекстах выборки. Полагается, что если два КС даются в описании одной и той же статьи, то они становятся связанными между собой через отнесение их к одному контексту. Соответствующим образом мы делаем вывод о связи между терминопольями, в которые входят указанные компоненты. С-карта автоматически генерируется на основе подсчета количества связей между полями в пределах всей выборки.

5.2. С-граф представляет собой визуальное представление С-карты.

### С-ГРАФЫ ТЕРМИНОВ И ТЕРМИНОПОЛЕЙ ПрО ЖУРНАЛА “ВОПРОСЫ КОГНИТИВНОЙ ЛИНГВИСТИКИ”

Генерация С-карты и С-графа может осуществляться на “сыром” материале: узлами структуры будут являться термины, а ребрами – показатели частоты их совместного использования в НКС.

Недостатком такой модели является ее размер: в нашем случае структурная модель состояла бы из 1650 узлов (общее количество КС в корпусе). Поэтому возможности использования С-карты и С-графа терминов ограничены и представляют интерес только в аспекте моделирования наиболее значимого фрагмента структуры ПрО. На рис. 1 представлен С-граф наиболее частотных терминов ПрО КОГНИТИВНАЯ ЛИНГВИСТИКА (порог значимости определялся как превышающий определенную частоту появления термина (в данном случае 0,01) в корпусе).

Представленная модель отражает “взгляд” на ПрО со стороны исследователей ее наиболее разработанного сектора. Однако данная модель передает лишь 8,1 % всех связей между терминами в ПрО. Поэтому для релевантного представления ПрО используется метод полевого анализа, результатом применения которого является структурированная в виде семантической карты (С-карты) и семантического графа (С-графа) *система терминопольей*. Кроме того, выявляются зависимости между временными (дата публикации), топологическими (географическими), социальными факторами и моделируемым информационным пространством ПрО. Соотнесенность семантики ПрО с территориальной, социальной, временной “картографией” позволяет понять логику развития ПрО, т.е. проследить пути, способы и механизмы реализации предметной интенции через социальные, географические и временные каналы обмена информацией (и их возможные комбинации).

Полевой анализ ключевых терминов к 446 научным статьям, опубликованным в журнале “Вопросы когнитивной лингвистики” за анализируемый шестилетний период позволил выделить 30 терминопольей (табл. 1).

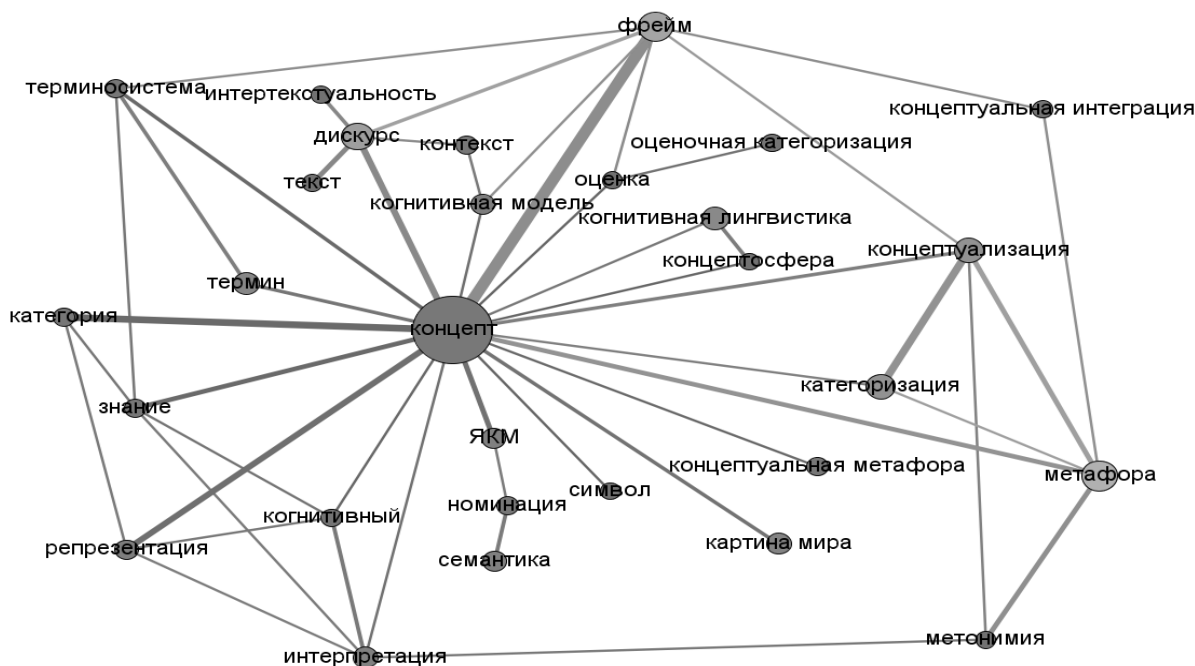


Рис. 1. Графосемантическая модель совместной встречаемости наиболее частотных терминов ПрО КОГНИТИВНАЯ ЛИНГВИСТИКА

*Примечание.* Размер узла отражает встречаемость КС в проанализированном материале. Толщина линии маркирует силу связи (частоту совместного использования в НКС) между терминами.

## Терминополья и частота их встречаемости в корпусе

Терминополье	Частота	Примеры терминов, входящих в терминополье
Структурная организация	0,633	организация, структура, фрейм, сценарий и мн. др.
Концепт и концептосфера	0,402	концепт, мифоконцепт, концепт дурак и мн. др.
Когниция	0,324	когнитивный, лингвокогнитивный уровень и мн. др.
Сфера культуры и лингвокультуры	0,316	лингвокультурема, мифоконцепт, идеология и мн. др.
Текст и дискурс	0,277	дискурс, интертекстуальность, дискурс-анализ и мн. др.
Психологические категории	0,250	память, экспрессия, ассоциации и мн. др.
Лексика и фразеология	0,245	лексика, идиомы, фразеологизм, термин и мн. др.
Коммуникация	0,215	манипуляция, коммуникативная стратегия и мн. др.
Понятийно-категориальная сфера	0,210	категория, модусные категории, категоризация и мн. др.
Семиотическое пространство	0,205	знак, семиосфера, биосемиотика и мн. др.
Язык	0,197	язык, языковое сознание и мн. др.
Грамматика	0,189	числительное, синтаксис, грамматикализация и мн. др.
Предметная область	0,189	экономика, корпоративная культура, политика и мн. др.
Семантика	0,162	семантические сети, семантическое поле и мн. др.
Тропы и фигуры	0,157	метафора, метонимия, антитеза, символ и мн. др.
Модель и моделирование	0,141	модель, когнитивное моделирование и мн. др.
Направления, школы	0,141	когнитивная лингвистика, теория дискурса и мн. др.
Методы исследования	0,136	шкалирование, сравнительный анализ и мн. др.
Процессуальность	0,130	динамичность, когнитивные процессы и мн. др.
Знание и информация	0,128	знание, информация, форматы знаний и мн. др.
Понимание и смысл	0,125	смысл, когнитивная модель переосмысления и мн. др.
Картина, образ, модель мира	0,117	картина мира, образ мира, модель мира и мн. др.
Речевая деятельность	0,109	речь, речекоммуникативный акт и мн. др.
Значение	0,101	перенос значения, значение, многозначность и мн. др.
Социосфера	0,088	гендер, социальная маркированность и мн.др.
Пространство и время	0,088	время, пространство, замкнутое пространство и мн.др.
Логические категории и операции	0,074	отрицание, противопоставление, аргумент и мн.др.
Сфера аксиологии	0,069	оценка, прагматический оценочный абсурд и мн.др.
Функциональность	0,059	функция, интерпретирующая функция и мн.др.
Языковая личность	0,040	языковая личность, авторская модальность и мн.др.

*Примечание:* Показатели частоты рассчитываются как отношение веса терминополья к количеству публикаций.

На рис. 2 представлено временное распределение частотности терминопольей за период 2008–2013 гг. График дает представление об общем интересе всех авторов журнала “Вопросы когнитивной лингвистики” к тем или иным частным научным направлениям, отражаемым в содержании каждого терминополья.

На рис. 2 видно, что СТРУКТУРНАЯ ОРГАНИЗАЦИЯ – наиболее значимое для ПрО терминополье, что свидетельствует о важности включения в модель

всех терминов ПрО, а не только ее наиболее разработанного сектора, в котором доминирующее положение отводилось КОНЦЕПТУ (см. рис. 1).

В то же время информационное пространство каждого конкретного исследования создается на основе взаимодействия нескольких терминопольей, что хорошо видно на представленном ниже рис. 3. В качестве примера проанализируем каждое КС из НКС процитированной выше статьи [15]:

ФОЛЬКЛОРНЫЙ ДИСКУРС относится к терминоплям КУЛЬТУРА И ЛИНГВОКУЛЬТУРА; ТЕКСТ И ДИСКУРС.

КОГНИТИВНО-ДИСКУРСИВНОЕ ИССЛЕДОВАНИЕ относится к терминоплям ТЕКСТ И ДИСКУРС; КОГНИЦИЯ; НАПРАВЛЕНИЯ, ШКОЛЫ.

КОММУНИКАТИВНЫЕ СТРАТЕГИИ относится к терминоплям СТРУКТУРНАЯ ОРГАНИЗАЦИЯ; КОММУНИКАЦИЯ.

ЦЕННОСТИ относится к терминоплю СФЕРА АКЦИОЛОГИИ.

ДИСКУРСООБРАЗУЮЩИЕ КОНЦЕПТЫ относится к терминоплям КОНЦЕПТ И КОНЦЕПТОСФЕРА; ТЕКСТ И ДИСКУРС.

Таким образом, в данном наборе ключевых слов актуализовано восемь разных терминоплей.

На рис. 3 видно, что минимальное количество полей, актуальных для каждой научной статьи – 1, а максимальное – 13; мода приходится на 6 полей. Данная модель наглядно показывает значимость исследования структурной связности терминоплей, формирующих информационное пространство как отдельной публикации, так и журнала в целом, так

как структурные связи отражают существующие в реальности комбинации (композиции) научных проблем и направлений.

Моделирование структуры терминоплей осуществляется с помощью метода графосемантического моделирования. С-граф терминоплей всего информационного пространства журнала “Вопросы когнитивной лингвистики” за шестилетний период 2008–2013 гг. представлен на рис. 4.

Ядро графа – терминопле СТРУКТУРНАЯ ОРГАНИЗАЦИЯ имеет связи почти со всеми терминоплями, репрезентируясь в НКС в соответствии с используемыми типами структур в том или ином терминопле. Например, соединение СТРУКТУРНОЙ ОРГАНИЗАЦИИ с КОНЦЕПТОМ И КОНЦЕПТОСФЕРОЙ частотно репрезентируется в терминах ФРЕЙМОВАЯ СТРУКТУРА КОНЦЕПТА, КОНЦЕПТУАЛЬНАЯ МЕЖФРЕЙМОВАЯ СЕТЬ и т.п. В то же время связь СТРУКТУРНОЙ ОРГАНИЗАЦИИ с ГРАММАТИКОЙ актуализует более традиционные типы структурных наименований: СТРУКТУРА НАРЕЧИЯ, АНАЛИТИЧЕСКИЙ ГЛАГОЛ, СТРУКТУРНАЯ СХЕМА ПРЕДЛОЖЕНИЯ и т.п.

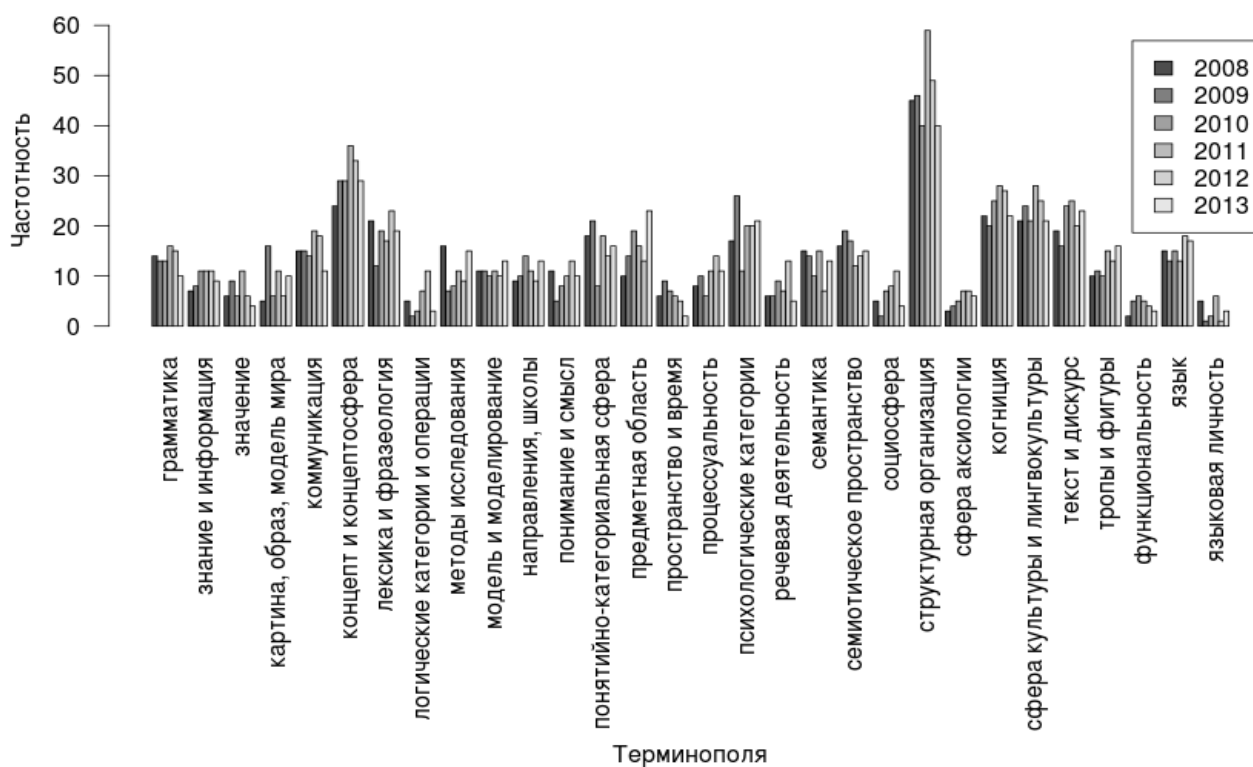


Рис. 2. Распределение частотности терминоплей за период 2008–2013 гг.

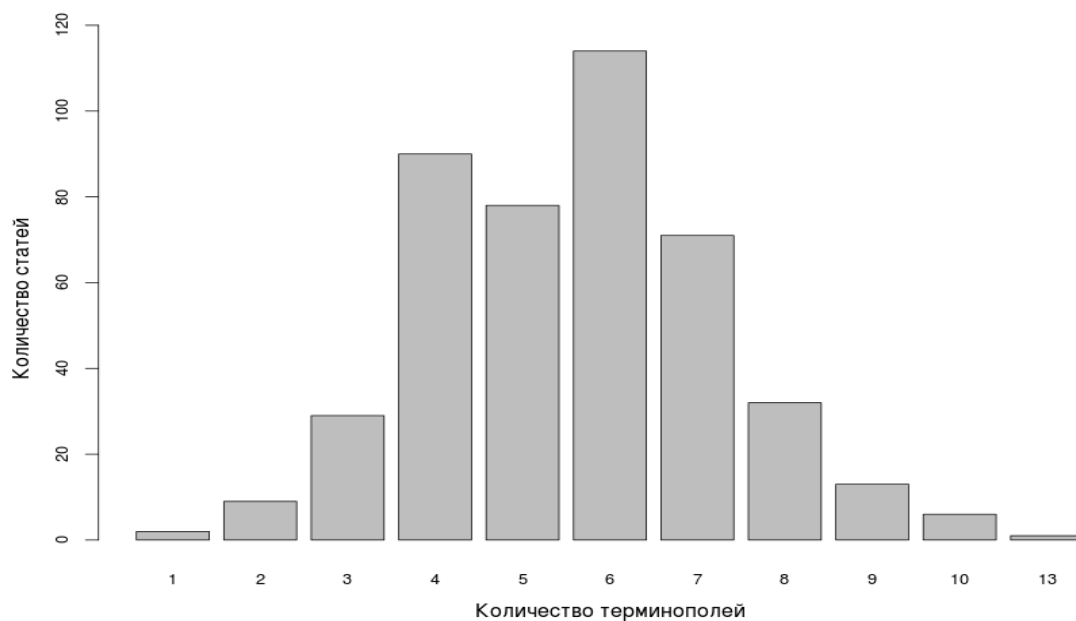


Рис. 3. Распределение количества терминоплей по научным статьям за период 2008–2013 гг.

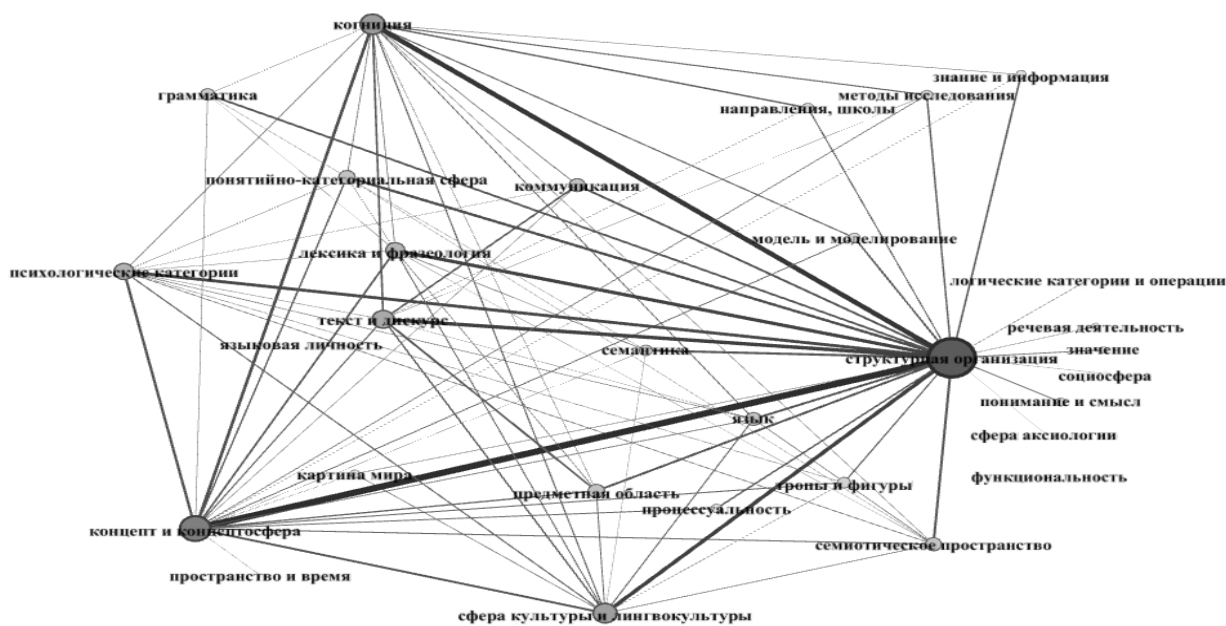


Рис. 4. С-граф терминоплей предметной области КОГНИТИВНАЯ ЛИНГВИСТИКА

*Примечание.* В С-графе вершины соответствуют терминоплям, а ребра – связям между ними. Размер вершины пропорционален частоте поля, а толщина ребра – силе связи между полями (частоте совместного присутствия двух терминоплей в разных НКС). Порог значимости для ребер превышает частоту 0,05.

На рис. 4 видно, что информационное пространство ПрО в журнале формируется за счет включения в лингвокогнитивную проблематику таких областей, как ТЕОРИЯ ДИСКУРСА и ЛИНГВОКУЛЬТУРОЛОГИЯ, при этом исследовательский “акцент” делается на изучении ЛЕКСИКИ и ФРАЗЕОЛОГИИ (ГРАММАТИКА отходит на второй план). Отметим также значимость терминопля ПСИХОЛОГИЧЕС-

КИЕ КАТЕГОРИИ, актуального для когнитивистики в целом, однако, отсутствие связей данного терминопля с важными для него терминоплями позволяет говорить о недостаточной разработанности психологической проблематики (в частности, отсутствуют связи с терминоплями ГРАММАТИКА, КАРТИНА МИРА, ЯЗЫКОВАЯ ЛИЧНОСТЬ и др.).

## ЧАСТНОНАУЧНАЯ ПРЕДМЕТНАЯ ОБЛАСТЬ КАК КОМПОЗИЦИЯ ТЕРМИНОПОЛЕЙ

**Частнонаучная предметная область** – представленная в виде математической модели композиция терминопольей, репрезентирующая отдельный научный сегмент в общей ПрО [16, с. 21]. В качестве метода, направленного на выделение частнонаучных ПрО из общей ПрО научного журнала, был использован *кластерный анализ*. Кластерный анализ производится над множеством контекстов, параметрами которых являются бинарные векторы, содержащие наборы полей соответствующих контекстов. Набором полей контекста называется множество терминопольей, связанных с данным контекстом посредством произвольного числа терминов (в нашем случае КС к публикациям). Поскольку каждому контексту соответствует частнонаучная ПрО, результат можно использовать в качестве оценки подобия публикаций и соответствующих им частнонаучных ПрО.

Выделение частнонаучных ПрО осуществляется с помощью метода нечеткой кластеризации C-means [17]. В отличие от других распространенных методов кластерного анализа, таких как K-means или самоор-

ганизуемые карты Кохонена, этот метод допускает принадлежность одного элемента двум и более кластерам, что в данной задаче позволяет отнести одну публикацию к нескольким направлениям в случае высокой неопределенности в ее описании. Кроме того, алгоритм метода C-means допускает возможность влияния на результат с помощью специального числового параметра  $m$ ,  $m \in \mathbb{R}$ ,  $m \geq 1$ . В качестве исходного значения обычно выбирается  $m=2$ , затем, в случае неудовлетворительного результата, выполняется его подстройка. Результаты данного исследования получены с базовым значением параметра  $m=2$ .

В результате кластеризации определяются: 1) частнонаучные ПрО, соотносимые с выделенными кластерами, 2) публикации, в той или иной мере соответствующие частнонаучным ПрО.

На основе полученных данных публикации были отсортированы по номеру кластера и расстоянию от его центра в порядке возрастания значений. В табл. 2 представлена кластеризация контекстов в виде 9-кластерной модели, которая была отобрана из спектра других моделей (состоящих из 6, 7, 8, 9 и 10 кластеров) на основании наибольшего соответствия ПрО КОГНИТИВНАЯ ЛИНГВИСТИКА.

Таблица 2

Частнонаучные предметные области (9-кластерная модель)

№ кластера	Композиции терминопольей	Условное обозначение	Количество статей	Репрезентативные статьи
1	1. Коммуникация 2. Структура 3. Текст и дискурс	ком-с-тд	66	[18]
2	1. Грамматика 2. Структура	гр-с	69	[19]
3	1. Концепт 2. Лексика и фразеология 3. Структура	к-лф-с	68	[20]
4	1. Направления 2. Структура 3. Когниция 4. Текст и дискурс	нш-с-кгн-тд	41	[21]
5	1. Знание и информация 2. Психол. категории 3. Структура	зн-псх-с	38	[22]
6	1. Концепт 2. Психол. категории 3. Структура	к-псх-с	45	[23]
7	1. Предметная область 2. Культура и лингвокультура 3. Текст и дискурс	по-клт-тд	43	[24]
8	1. Концепт 2. Лексика 3. Предметная область 4. Семиотика 5. Структура	к-лф-по-сп-с	26	[25]
9	1. Структура 2. Культура и лингвокультура 3. Язык	с-клт-яз	50	[26]



Выявленные на основе кластеризации публикаций частнонаучные ПрО представляют собой основные направления исследований в 2008–2013 гг., а представленная на рис. 5 гистограмма распределения выделенных кластеров (частнонаучных ПрО) по временным срезам отражает временную динамику научных интересов авторов журнала. Представленная диаграмма дает возможность оценить, насколько динамичны, изменчивы и в то же время в отдельных аспектах взаимосвязаны области научных интересов журнала.

Так, в частности, увеличение в последние годы количества публикаций кластера 3 “к-лф-с” (реконструкция концептов на основе лексического анализа) свидетельствует о выборе журналом в качестве приоритетных проблем концептологии. Альтернативное направление (кластер 6 “к-псх-с”), имеющее в 2008 г. близкую стартовую позицию, в дальнейшем вытесняется на периферию (особенно это заметно по количественным данным 2013 г.).

Полагаем, что вытеснению данного направления способствовали набирающие популярность исследования в рамках кластера 8 (к-лф-по-сп-с), связанные с исследованием терминосистем и концептосфер специального знания. Аналогичным образом связаны между собой исследования в области теории дискурса (кластеры 1, 4 и 7) и лингвокультурологии (кластеры 7 и 9).

На рис. 6 представлена карта научных интересов авторов журнала (сгруппированных по географической принадлежности), составленная в соответствии с выделенными частнонаучными ПрО.

Рис. 6 дает наглядное представление, с одной стороны, о научных интересах авторов, а с другой – об обоснованности самой группировки по городам, показывающей “научную географическую специализацию”. В том случае, когда речь идет о ведущих научных центрах (Москва, Санкт-Петербург) и достаточно большом объеме публикаций, сферы научных интересов распределяются почти равномерно. Небольшие научные центры даже при внушительном объеме публикаций имеют тенденцию к специализации. Особенно это заметно с публикациями исследователей из Махачкалы (почти все находятся в кластере 2) и статьями представителей Тверской лингвистической школы.

В то же время изучение научной “географической специализации” нужно осуществлять с временной “привязкой” публикаций, поскольку это позволит: 1) сопоставить синхронные изменения в информационном пространстве ПрО (на рис. 6 сравниваются работы, созданные в разные периоды времени); 2) выявить наличие/отсутствие изменений в исследовательских интересах к тем или иным научным проблемам; 3) установить зависимости в появлении интересов к тем или иным научным проблемам.

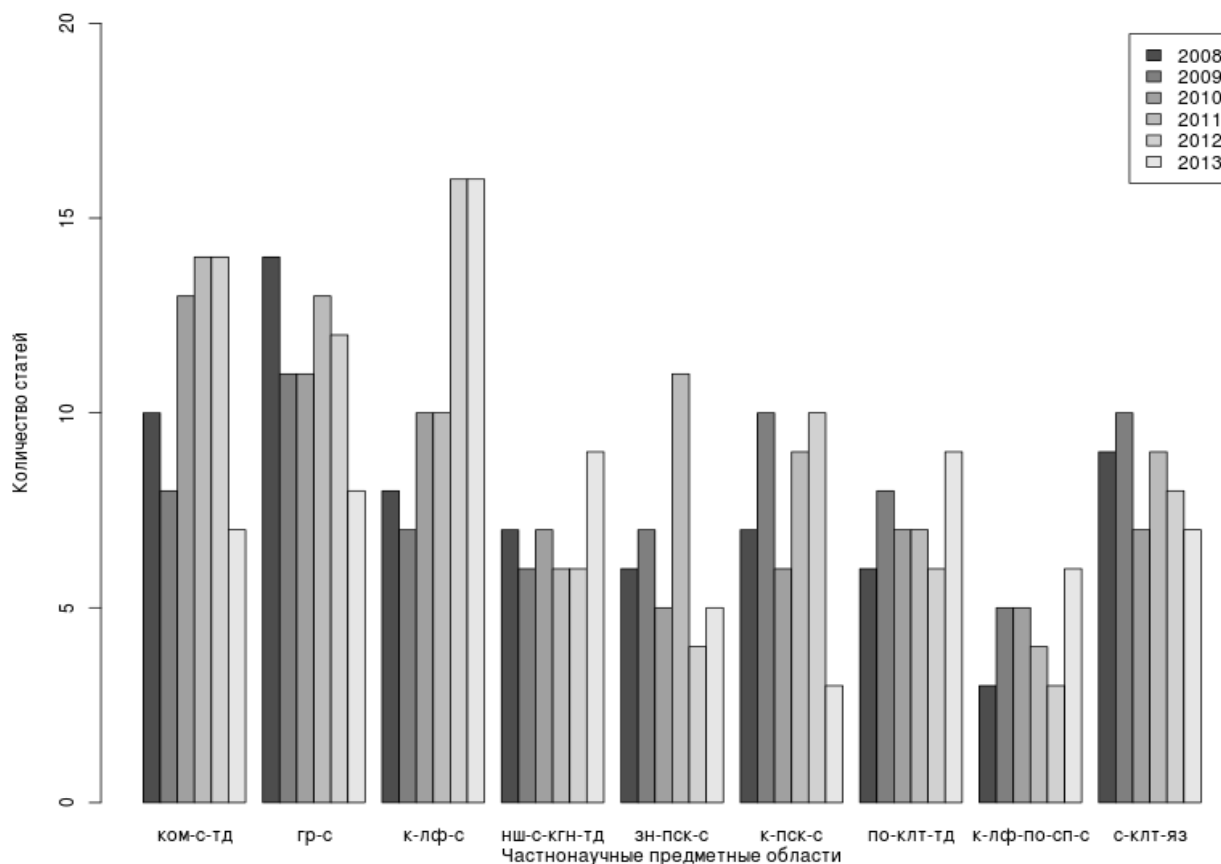


Рис. 5. Распределение выделенных кластеров (частнонаучных предметных областей) по временным срезам 2008–2013 гг.

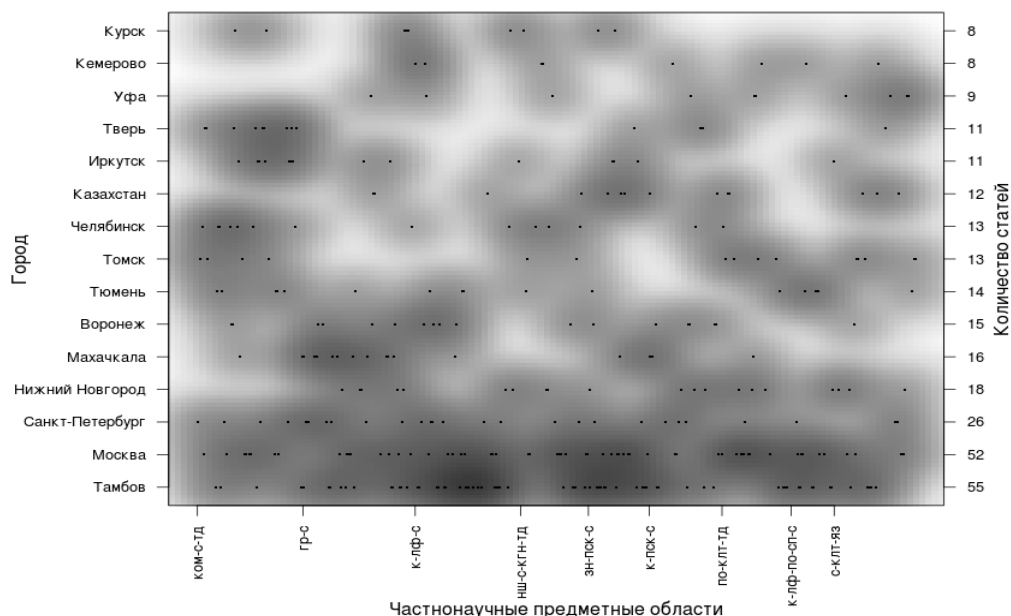


Рис. 6. Карта научных интересов авторов журнала “Вопросы когнитивной лингвистики” за период 2008–2013 гг. (по частнонаучным предметным областям)

*Примечание.* Точками обозначены отдельные публикации. Центры кластеров располагаются на линиях, проведенных вертикально из намеченных кластеров (над их названиями). Чем ближе расположена точка к такой вертикальной линии, тем более репрезентативна для данного кластера публикация.

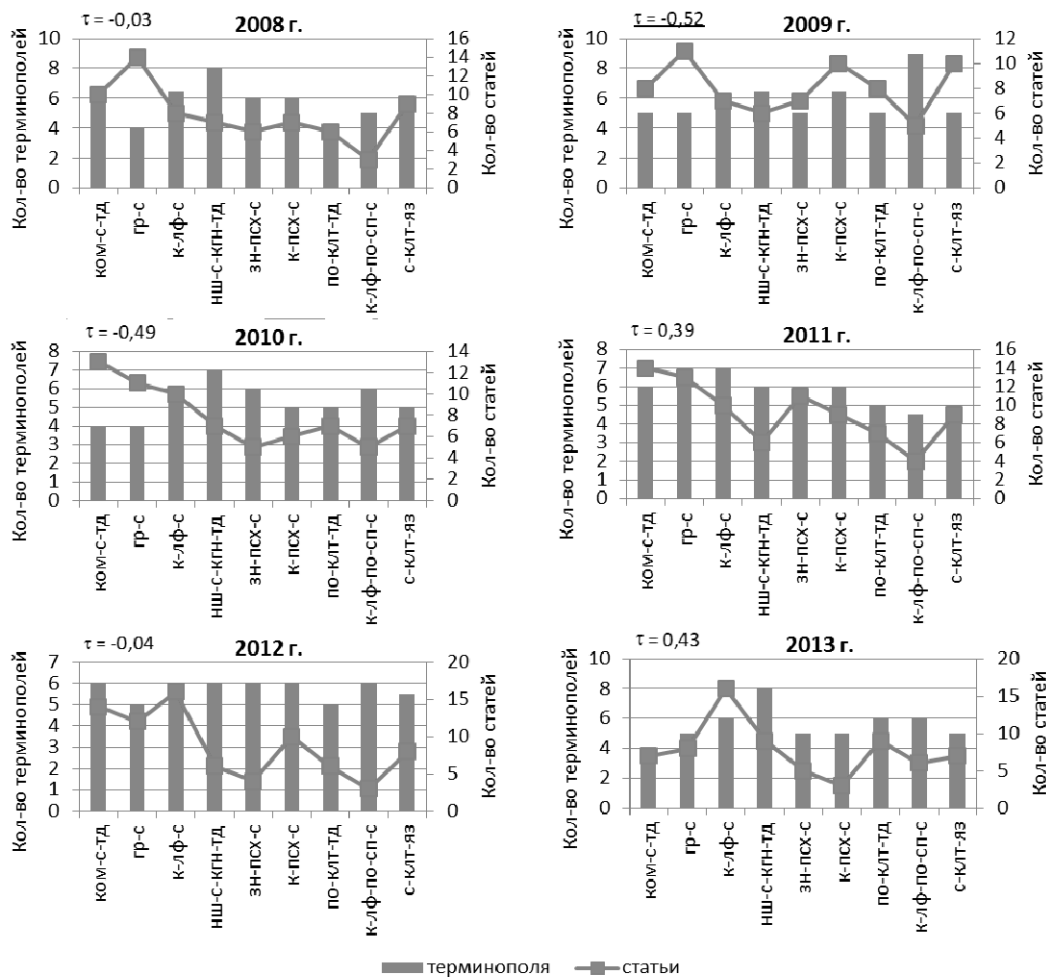


Рис. 7. Временная динамика распределения по кластерам среднего количества терминопольей в НКС к публикациям и количества публикаций (группировка по годам).

*Примечание.* Подчеркнуты показатели значимых корреляций при  $p < 0,05$ .

Имеющиеся данные позволяют рассматривать проблему в следующем аспекте: 1) по каждой частнонаучной ПрО есть данные, свидетельствующие об интересе к ней в любом временном срезе (см. рис. 5); 2) в то же время имеются данные о среднем количестве терминопольей, актуализованных в НКС к статьям каждой частнонаучной ПрО в тех же временных срезах. Можно предположить, что разработка частнонаучной ПрО будет приводить к ее усложнению, что на формальном уровне может выражаться в увеличении среднего количества терминопольей в НКС к статье. А временная динамика среднего количества терминопольей в НКС к статье в частнонаучной ПрО должна соотноситься с количеством публикаций по данной тематике (т.е. с интересом к изучаемому проблемному полю).

На рис. 7 и 8 соотнесены два ряда данных – среднее количество терминопольей в НКС к статье и количество публикаций. Рис. 7 представляет данные, сгруппированные по годам, рис. 8 – по выделенным кластерам. На каждом рисунке отображаются показатели коэффициента корреляции Кендалла (показатели значимых корреляций при  $p < 0,05$  подчеркнуты).

На рис. 8 видно, что только три частнонаучные ПрО имеют корреляцию уровня сложности ПрО с ее популярностью. Однако со временем корреляции между сложностью и популярностью увеличиваются (см. рис. 7): от отсутствия корреляции (2008 г.) до отрицательной корреляции (2009 и 2010 гг.) и изменения знака на положительную корреляцию (2011 и

2013 гг.). Обнаруженные закономерности между уровнем сложности/разработанности ПрО и ее популярностью среди исследователей нуждаются в более масштабной проверке и в случае подтверждения могут использоваться для оценки, прогноза и планирования редакционно-издательской деятельности научных журналов или для дополнительного обоснования требуемого уровня поддержки научными фондами тех или иных исследовательских направлений (о существующих методах см., в частности, [27]).

Для прогнозирования значений количества терминопольей и публикаций в 2014 и 2015 гг. воспользуемся методом Multiple Regression (Множественная регрессия), реализованном в статистическом пакете Statistica 8.0 (см., например, [28, с. 180–190]). Прогнозируемые показатели рассчитываются для каждого выделенного кластера; результаты представлены на рис. 8. На рис. видно, что прогнозируется наибольший рост количества публикаций кластера 3 “к-лф-с” и уменьшение количества публикаций конкурирующих с ним кластеров 6 “к-псх-с” и 8 “к-лф-по-сп-с”. Между тем, кластер 3 – наиболее простая частнонаучная ПрО (достаточно сказать, что она значительно чаще остальных используется в квалификационных работах). Кластеры 6 и 8 представляют, на наш взгляд, большую теоретическую и прикладную ценность, поэтому складывающаяся ситуация требует некоторого “выравнивания” частнонаучных ПрО, придания импульса наиболее значимым из них.

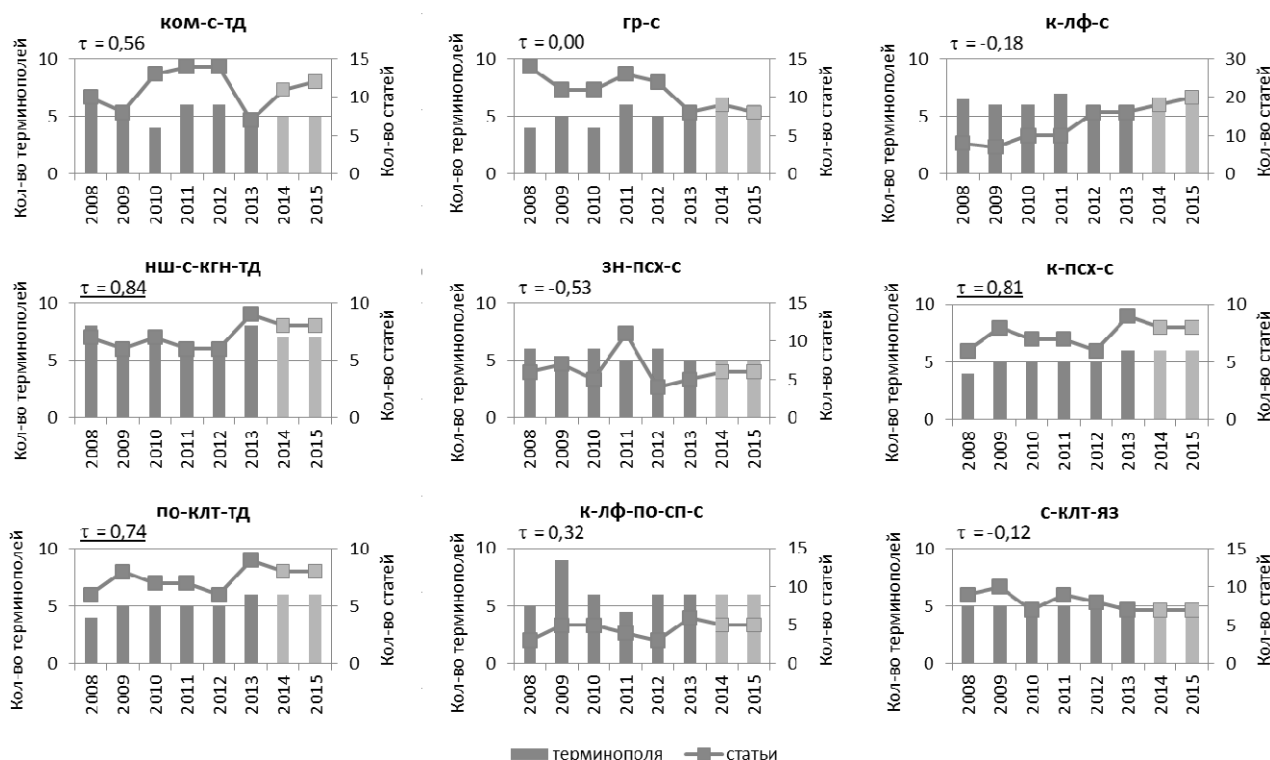


Рис. 8. Временная динамика распределения (с прогнозом) по кластерам среднего количества терминопольей в НКС к публикациям и количества публикаций (группировка по кластерам)

*Примечание.* Подчеркнуты показатели значимых корреляций при  $p < 0,05$  (прогнозируемые значения не учитывались).

## ВЫВОДЫ

Изучение репрезентации ПрО в тематическом научном журнале представляет интерес в самых разных аспектах. Анализ ПрО с позиции “живой науки” позволяет выявить частнонаучные ПрО, определить их “веса” и динамику изменений, а также представить прогнозные модели развития. Представленные результаты демонстрируют возможности используемого метода исследования и позволяют говорить о концептосфере научного журнала как о постепенно оформляющейся системе терминопольей; при этом сам процесс оформления концептосферы (ПрО) в отдельных ее секторах имеет разную динамику. Отчасти данная динамика обусловлена социальными факторами, в частности, научным статусом участников (можно отметить более высокую корреляцию между динамикой сложности частнонаучной ПрО и количеством публикаций в случае большего участия в ее разработке докторов наук). Вероятно, влияние на динамику отдельных частнонаучных ПрО оказывают также широта географического распространения, традиции научных школ, возможности междисциплинарного синтеза.

Полагаем, что существующие методы и теоретический базис терминоведения, рассмотренные в контексте современных информационных технологий и методов математического моделирования, позволяют решать не только классификационные задачи, но и задачи оценки, прогнозирования и управления исследовательской деятельностью, в том числе и в рамках создания систем поддержки принятия решений в данной сфере.

## СПИСОК ЛИТЕРАТУРЫ

1. Nederhof A.J., van Leeuwen T.N., van Raan A.F.J. Highly cited non-journal publications in political science, economics and psychology: a first exploration // *Scientometrics*. – 2010. – Vol. 83, Iss. 2. – P. 363–374.
2. Григорьев В.К. Научные журналы – механизм объективной оценки грантополучателей в технологии закрепления вузовской молодежи в науке // *Дистанционное и виртуальное обучение*. – 2013. – № 4. – С. 110–119.
3. Álvarez-de-Toledo-Saavedra L. Bibliographic control and dissemination of the University of Oviedo scientific output // *El profesional de la información*. – 2012. – Vol. 21, Iss. 6. – P. 639–642.
4. Стародубов В.И., Куракова Н.Г., Цветкова Л.А., Арефьев П.Г., Кураков Ф.А. Проблемы оценки мирового уровня конкурентоспособности российской науки на примере национальной клинической медицины // *Научно-техническая информация*. Сер.1. – 2012. – № 8. – С. 1–15; Starodubov V.I., Kurakova N.G., Tsvetkova L.A., Aref'ev P.G., Kurakov A.F. The Problems Associated with the Evaluation of World-Class Competitiveness of Russia's Science, as Illustrated by Clinical Medicine // *Scientific and Technical Information Processing*. – 2012. – Vol. 39, № 3. – P. 139–152.
5. Москалева О.В. Можно ли оценивать труд ученых по библиометрическим показателям? // *Управление большими системами. Специальный выпуск 44: Наукометрия и экспертиза в управлении наукой*. – 2013. – С. 308–331.
6. Базы данных ИНИОН. – URL: <http://www.inion.ru/index6.php> (дата обращения: 18.04.2014).
7. Мдивани Р.Р. Тезаурусы ИНИОН РАН по социальным и гуманитарным наукам // *Научно-техническая информация*. Сер. 1. – 2013. – № 7. – С. 23–27.
8. Arano S. Thesauruses and ontologies // *Hiper-text.net*. № 3. – URL: <http://eprints.rclis.org/8972/2/12.pdf> (дата обращения: 18.04.2014).
9. Wielinga B.J., Schreiber A.Th., Wielemaker J., Sandberg J.A.C. From thesaurus to ontology // *K-CAP '01: Proceedings of the 1st international conference on Knowledge capture*. – ACM New York, 2001. – P. 194–201.
10. Гладун А.Я., Рогушина Ю.В. Использование онтологических знаний и тезаурусов для объективного профилирования специалистов // *Искусственный интеллект*. – 2006. – № 3. – С. 379–390.
11. Никитина С.Е. Семантический анализ языка науки. На материале лингвистики. – М.: Наука, 1987. – 146 с.
12. Кузнецов А.М. От компонентного анализа к компонентному синтезу. – М.: Наука, 1986. – 126 с.
13. Морозова Л.А. Терминознание: Основы и методы. – М.: ГНО “Прометей” МПГУ, 2004. – 144 с.
14. Шур Г.С. Теории поля в лингвистике. – М.: Эдиториал УРСС, 2009. – 264 с.
15. Эмер Ю.А. Фольклорный дискурс: когнитивно-дискурсивное исследование // *Вопросы когнитивной лингвистики*. – 2011. – № 2. – С. 50–60.
16. Белоусов К.И., Баранов Д.А., Зелянская Н.Л. Научный коллектив и его предметные области (к вопросу о методах эффективного планирования научной деятельности) // *Научно-техническая информация*. Сер. 1. – 2014. – № 4. – С. 13–26.
17. Kaymak U., Setnes M. Extended fuzzy clustering algorithms // *Erasmus Research Institute of Management*. – 2000. – URL: <http://repub.eur.nl/pub/57/erimrs20001123094510.pdf> (дата обращения: 18.04.2014).
18. Виноградова С.А. Инструменты речевой манипуляции в политическом медиадискурсе // *Вопросы когнитивной лингвистики*. – 2010. – № 2. – С. 95–101.
19. Березина О.А. К вопросу о языковой репрезентации категории безличности // *Вопросы когнитивной лингвистики*. – 2010. – № 4. – С. 39–51.
20. Непрокина Ю.А., Пискунова С.В. Концепт ЛЕТО и его воплощение в языковой культуре тамбовского края // *Вопросы когнитивной лингвистики*. – 2009. – № 4. – С. 123–128.

21. Белоусов К.И., Зелянская Н.Л. Лингвосомиотическое моделирование обыденной географической картины мира // Вопросы когнитивной лингвистики. – 2013. – № 2. – С. 73–85.
22. Новикова М.Г. Возможность графической интерпретации процесса понимания // Вопросы когнитивной лингвистики. – 2011. – № 3. – С. 132–137.
23. Коваленко Г.Ф. Межтекстовое ассоциативно-смысловое поле как способ представления художественного концепта LOVE в идиостиле И. Стоуна // Вопросы когнитивной лингвистики. – 2012. – № 3. – С. 76–83.
24. Пеньков Б.В. Ядро и периферия образовательного дискурса (на материале британской средней школы) // Вопросы когнитивной лингвистики. – 2010. – № 3. – С. 84–91.
25. Зыкова И.В. Семиотика музыки в построении фразеологического значения (лингвокультурологический подход) // Вопросы когнитивной лингвистики. – 2012. – № 4. – С. 108–117.
26. Тахтарова С.С. Концепт TOLERANZ в немецкой лингвокультуре // Вопросы когнитивной лингвистики. – 2008. – № 1. – С. 64–71.
27. Рудцкая Е.Р., Хрусталева Е.Ю., Цыганов С.А. Методы накопления научного знания для инновационного развития российской экономики (опыт РФФИ) // Проблемы прогнозирования. – 2009. – № 3. – С. 134–139.
28. Вуколов Э.А. Практикум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel. – М.: ФОРУМ, 2008. – 464 с.

*Материал поступил в редакцию 09.07.14.*

## **Сведения об авторах**

**БЕЛОУСОВ Константин Игоревич** – доктор филологических наук, профессор кафедры теоретического и прикладного языкознания Пермского государственного национального исследовательского университета  
e-mail: belousovki@gmail.com

**БАРАНОВ Дмитрий Александрович** – аспирант кафедры математического обеспечения информационных систем Оренбургского государственного университета  
e-mail: baranov@semograph.com

**ЕРОФЕЕВА Елена Валентиновна** – доктор филологических наук, профессор кафедры теоретического и прикладного языкознания Пермского государственного национального исследовательского университета  
E-mail: elevaer@gmail.com

**ЗЕЛЯНСКАЯ Наталья Львовна** – кандидат филологических наук, ведущий научный сотрудник кафедры теоретического и прикладного языкознания Пермского государственного национального исследовательского университета  
E-mail: zelyanskaya@gmail.com

**ИЧКИНЕЕВА Дилара Ахметовна** – кандидат филологических наук, доцент кафедры иностранных языков естественных факультетов Башкирского государственного университета  
E-mail: dilaraichkineeva@gmail.com