

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК [005.551 : 005.552.3] : 001.89

К.И. Белоусов, Д.А. Баранов, Н.Л. Зелянская

Научный коллектив и его предметные области (К вопросу о методах эффективного планирования научной деятельности)*

Описывается исследовательская программа моделирования информационного пространства, сложившегося в отдельной частнонаучной предметной области исследовательского коллектива, включающая выделение терминов, группировку их в терминополь, изучение временной динамики актуализации терминопольей. На основе алгоритма нечеткой кластеризации C-means предлагается кластерная модель исследовательских направлений научного коллектива и карты его научных интересов. Представленные модели могут быть востребованы для планирования деятельности участников научного коллектива.

Ключевые слова: научное исследование, научная карта, предметная область, прогнозирование, графосемантическое моделирование, статистические методы, информационная система “Семограф”

ВВЕДЕНИЕ

Оценка эффективности деятельности индивидуальных и коллективных агентов научного производства обычно осуществляется на основе наукометрических показателей, относящихся к публикационной активности. Сбором информации и ее обобщением занимаются международные институты (Institute for Scientific Information (ISI), International Mathematical Union (IMU), International Council for Industrial and Applied Mathematics (ICAM)), созданы реферативные / библиографические базы данных научных публикаций, крупнейшие и наиболее авторитетные из которых Web of Science и Scopus. Наукометрические проблемы интересуют и крупнейшую поисковую систему Google, разработавшую сервис Google Scholar. Российский аналог «Научная электронная библиотека» (ELIBRARY.ru) предложила национальный российский индекс научного цитирования – РИНЦ. Основу расчетов разнообразных индексов (цитируемости, самоцитируемости, Хирша, импакт-фактора и многих других) составляют количественные показатели числа публикаций и цитирований за определенные периоды времени. Многочисленные индексы / агрегаторы научного контента (в том числе и отраслевые) поставляют данные для исследований в области наукометрии, особенно популярной

за рубежом. Исследователи оценивают возможность применения количества цитирований, вводят индексы Hirsch h-index, Hirsch h-core, g-index, R-index; AR-index, Jin A-index для определения эффективности научной деятельности, тестируют существующие индексы научного контента (WoS, Scopus, Google Scholar и др.), исследуют особенности оценки эффективности в гуманитарных, социальных, естественнонаучных и технических направлениях [1-5], изучают возрастные и гендерные аспекты эффективности и производительности в науке [6-8], осуществляют анализ междисциплинарных взаимодействий [9, 10] и возможности использования индексов для сопоставления национальных наук и научных специализаций [11-14].

Эти же формальные показатели используются как основа планирования научной деятельности, в том числе, в аналитических инструментах InCites и SciValSpotlight (приложения, работающие на базе наукометрических показателей WoS и Scopus), созданных как инструмент принятия решений в области финансирования проектов и научной деятельности коллективов [15; 16].

Формализованный подход к оценке эффективности научной деятельности имеет неоспоримые плюсы. В частности, он позволяет делать мгновенные “срезы” состояния научных исследований и планировать поддержку отдельных ученых или научных коллективов. Однако индексы, вычисляемые на основе публикационной активности и цитируемости,

* Исследование выполнялось при финансовой поддержке Российского гуманитарного научного фонда (проект № 12-34-01087).

дают показатели эффективности *результатов* научной деятельности, но не могут быть положены в основу планирования своей работы самими исследователями.

МЕТОДОЛОГИЯ

В свете изложенного становятся актуальными поиски новых форм, методов, инструментов эффективного планирования научной деятельности, в первую очередь, таких, которые бы основывались на состоянии информационного пространства, сложившегося в отдельной научной предметной области (далее – ПрО, частнонаучной ПрО) с возможностью обобщения до информационных границ науки в целом.

Методы тезаурусного моделирования и полевого анализа частнонаучной предметной области

Базовым методом исследования информационного пространства частнонаучной предметной области является широко используемый в онтологическом инжиниринге и терминоведении *метод построения тезаурусов* [17-21]. Тезаурус представляет собой модель языка науки, в которой связь между понятиями осуществляется “горизонтальными” и “вертикальными” семантическими типами отношений. Иерархическая структура тезауруса позволяет генерировать модели как максимально отдаленные от конкретного материала (т.е. от терминов, используемых в научных статьях), так и приближенные к нему, передающие специфику предмета отдельного научного исследования. Между тем использование тезаурусного подхода для представления ПрО накладывает определенные ограничения на создаваемую модель, наиболее спорным из которых является однозначность классификации – отнесения понятия к определенному узлу тезауруса. Тезаурусный подход становится нерелевантным моделируемой действительности, когда речь заходит о структурировании понятий ПрО с неоднозначными семантическими отношениями. Поэтому в лингвистике тезаурусное моделирование чаще применяется утилитарно, для составления языковых тезаурусов, а как метод структурирования единиц языковой реальности повсеместно, в том числе в терминоведении [22], используется *полевой анализ* [23, 24], результатом применения которого являются семантические (понятийные, терминологические, тематические, ассоциативные и др.) поля. Преимуществами полевого анализа перед тезаурусным моделированием являются: 1) возможность существования любых семантических отношений между единицами поля, так как для отнесения единицы к полю достаточно наличия одной общей семы; 2) возможность отнесения одной и той же единицы к нескольким полям. В то же время структура семантического поля более “размыта”, чем структура тезауруса. Поэтому мы полагаем, что моделирование частнонаучных ПрО должно основываться на полевом анализе с приданием терминоплям иерархической структуры.

К настоящему времени сформирован инструментарий онтологического моделирования ПрО, воплощенный в ряде редакторов для создания тезаурусов и

онтологий (WordNet, РузТез, Сус / ResearchСус, Protégé, OilEd, InTez, OntoGrid и др.). В нашей работе используется информационная система графосемантического моделирования – ИС “Семограф” [25], которая предназначена для извлечения знаний о ПрО из информационных массивов [26, 27]. Отличие ИС “Семограф” от существующих онто- и тезаурусных редакторов в том, что она основывается на построении полевых моделей, внутри которых может устанавливаться та или иная иерархия, т.е. предполагает интеллектуальное движение “от материала” при оформлении границ терминопольей.

Возможности создаваемых моделей частнонаучных предметных областей

В ИС “Семограф” осуществляется структурирование множества понятий исследуемой ПрО и контекстов использования терминов и терминопольей, результатом чего становится выявление ведущих и перспективных направлений исследований, обнаружение наиболее репрезентативных публикаций в интересующей научной области, анализ и прогнозирование состояния ПрО (в самых разнообразных ее “разрезах” (временном, пространственном, социальном, концептуальном и др.)), оценка эффективности исследовательской деятельности агентов (индивидуальных и коллективных) научного производства, подбор близких по научным интересам изданий для публикации результатов научной деятельности.

В данном контексте под *эффективным планированием* деятельности агента научного производства мы понимаем такое планирование научной деятельности, при котором агент, сообразно своим научным интересам и достижениям, на основании математических моделей развития научной области отбирает частнонаучные ПрО, которые на момент их выбора набирают “вес”.

Эффективность выбора ПрО обусловлена факторами: 1) содержанием и структурой собственной ПрО агента (микро-ПрО); 2) динамикой роста ПрО научного направления (макро-ПрО) – в том числе и в виде прогноза и 3) временным вхождением агента в данную ПрО (когда? на каком этапе?). Очевидно, что способность предвидеть состояние науки в той или иной временной перспективе является важнейшей составляющей успешности.

Для эффективного планирования научной деятельности агенту необходимо соотносить разрабатываемую им микро-ПрО с макро-ПрО, что можно представить в виде следующей исследовательской программы:

- 1) построение на основе собственных публикаций модели разрабатываемой микро-ПрО и выбор доминантных направлений;
- 2) создание корпуса текстов макро-ПрО по доминантным направлениям и построение на основе корпуса модели ПрО научного направления;
- 3) анализ макро-ПрО в аспекте временной динамики, построение математических моделей развития ПрО, выявление наиболее перспективных направлений;
- 4) соотнесение структуры микро-ПрО с перспективными направлениями и выработка решения о корректировке микро-ПрО.

В настоящей статье представим реализацию первого этапа предложенной исследовательской программы на примере научной деятельности коллектива лаборатории прикладных и экспериментальных лингвистических исследований Пермской социопсихолингвистической школы¹ за трехлетний период 2010-2012 гг.

Сбор и первичный анализ материала в ИС “Семограф”

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

В качестве *материала* исследования рассматривались научные статьи исследовательского коллектива, опубликованные в рецензируемых журналах (по данным <http://elibrary.ru>)². *Операциональными единицами* послужили базовые термины, выносимые авторами публикаций в наборы ключевых слов (далее – НКС) к своим статьям. Обращение к ключевым словам научных публикаций обусловлено тем, что, во-первых, посредством ключевых слов авторы сами обозначают доминанты концептуального пространства своих исследований; во-вторых, ключевые слова научных публикаций представляют собой легко формализуемый конструкт в рамках большого корпуса текстов; в-третьих, ключевые слова, в отличие от статьи, к которой они относятся, обычно доступны для автоматизированного извлечения.

Работа в ИС “Семограф” включает создание Проекта (рабочего пространства, в котором осуществляется исследовательский цикл). Фрейм Проекта с учетом специфики данного исследования состоит из следующих элементов.

1. Контекст – в нашем случае, описание научной статьи. Например: Ерофеева Е.В., Ерофеева Т.И. Человек и текст: антропоцентрический подход к исследованию // Вестник Пермского университета. Российская и зарубежная филология. – 2010. – № 4. – С. 28-33 [28].

2. Вся выборка контекстов (описание всех статей научного коллектива).

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

3.1. Метаданные (тип текстовых данных, позволяющих ввести дополнительную информацию о статье: автор / авторы, название статьи, год публикации (2010, 2011, 2012), название журнала, в котором издана статья).

3.2. Семантические компоненты (наборы ключевых слов к каждой статье). Например, к цитируемой статье авторы дают следующий НКС: спонтанная речь; социальные факторы; синтаксис; предложение; синтагма; фраза; статистическое моделирование [28].

3.3. Семантические поля – в нашем случае, терминополья, объединяющие отдельные научные понятия (выраженные с помощью КС). Терминополье представляет в модели входящие в него термины, что позволяет решить проблему вариативности ключевых слов (в том числе и терминологических повторов разного рода, например, синонимических). Для выполнения данного этапа требуется привлечение экспертов моделируемой научной отрасли. В задачу экспертов входит группировка доступных терминов (ключевых слов к статьям) в терминополья, формулировка названий терминопольей³. 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.

4. 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 С-карта и С-граф:

4.1. Семантическая карта (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.

4.2. Семантический граф (С-граф) представляет собой графическую экспликацию связей между выделенными терминопольями в С-карте.

МОДЕЛИРОВАНИЕ НАУЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ ИССЛЕДОВАТЕЛЬСКОГО КОЛЛЕКТИВА

Анализ ключевых терминов к 28 научным статьям исследовательского коллектива за анализируемый трехлетний период позволил выделить 21 терминополье (табл. 1).

¹ Границы научного коллектива устанавливались на основе принадлежности исследователя к коллективу лаборатории прикладных и экспериментальных лингвистических исследований Пермского государственного национального исследовательского университета (ПГНИУ). Авторы выражают признательность проф. Е.В. Ерофеевой, руководителю лаборатории прикладных и экспериментальных лингвистических исследований Пермской социопсихолингвистической школы, за консультации по возникшим в ходе исследования проблемам анализа научных направлений Школы.

² Общий объем публикаций научного коллектива за трехлетний период составил более 100 публикаций, из которых в рецензируемых журналах было издано 28 научных статей.

³ Допускаем, что классификация одного и того же научного материала разными экспертами может привести к несовпадающим результатам (количеству терминопольей, их содержанию или наименованию). Однако этот “недостаток” является скорее достоинством модели, так как будет показывать жизнеспособность научного взгляда в созданной на его основе модели ПрО.

Частотность обращения к терминопоям во всех публикациях научного коллектива

Терминополье	Частотность*	Термины, входящие в терминополье
Социокультурные аспекты функционирования языка	0,72	субкультура, гендерные стереотипы, ценностные установки, этнические ценности, групповая идентичность, ролевика, социальная группа, социальная база и др.
Методы исследования	0,48	ассоциативный эксперимент, стратификационное моделирование, статистическое моделирование, дисперсионный факторный анализ, социолингвистический метод и др.
Социально-территориальная дифференциация языка (С-ТДЯ)	0,41	профессионализм, идиолект, локализм, социолект, сленг, территориальное варьирование языка, городские социолекты, социальная дифференциация языка и др.
Речь	0,34	письменная речь, речевой жанр, городская речь, спонтанная речь, дифференциация речи и др.
Теория социолингвистики	0,31	социолингвистика, социальная диалектология, социопсихолингвистический аспект, микросоциолингвистика и др.
Лексика	0,28	термин, лексика, человек, заимствование, лексико-семантическое поле и др.
Ментальный лексикон	0,28	ментальный лексикон, актуальный лексикон, внутренний лексикон, лексикон и др.
Семантика	0,28	тематическая группа, лексико-семантическое поле, актуализация, (тематические) группы, подгруппы и др.
Язык	0,24	языковое сознание, языковые структуры, языковая идентичность, гендерные языковые стереотипы и др.
Речь города	0,21	локализм, речь города, городские социолекты, городская речь
Когниция	0,21	языковое сознание, картина мира, когнитивные структуры, когнитивные способности и др.
Структуры	0,21	тематическая группа, когнитивные структуры, лексико-семантическое поле, структура, группы, подгруппы
Гендер	0,17	гендер, гендерные стереотипы, гендерные языковые стереотипы и др.
Языковые процессы	0,17	механизм, интеграция (языковых) факторов
Грамматика	0,14	глагол, категория глагольного вида, конкуренция видов, синтаксис и др.
Языковая личность	0,14	речевое поведение, идиолект, социальный опыт говорящего
Коммуникация	0,14	коммуникативная ситуация, речевая компетенция, церковно-религиозная сфера коммуникации и др.
Языковые контакты	0,10	заимствование, билингвизм, билингвы-татары, языковой контакт
Фонетика	0,07	синтагма, фраза, фонетика, фонетические особенности
Художественные средства	0,03	поэтическая энергетика
Текст	0,03	текст

* Показатели частотности рассчитываются как отношение веса терминопоя к количеству публикаций.

Временное распределение частотности терминопоя за период 2010-2012 гг. представлено на рис. 1, который дает представление об общем интересе всех участников научного коллектива к тем или иным частным научным направлениям, отражаемым в содержании каждого терминопоя.

Например, можно отметить, что в 2012 г. усиление интереса к проблемам языка (при стабильно высоком внимании к исследованиям речи) сопровождается ростом публикаций по проблематике языковых процессов и языковой личности. В то же время, несмотря на наглядную демонстрацию соотношения отдельных терминопоя друг с другом (это, кроме уже

названных, *социокультурные аспекты функционирования языка, гендер, социально-территориальная дифференциация языка, речь города* и некоторые другие), рис. 1 все же не дает представления о связности научных проблем в исследованиях (например, понимания того, что лексика интересует исследователей, в первую очередь, как инструмент изучения ментального лексикона или социально-территориальной дифференциации языка).

Понятийное пространство каждого конкретного исследования создается на основе взаимодействия нескольких терминопоя, что хорошо показано на рис. 2. В качестве примера анализируем каждое

ключевое слово из набора ключевых слов к цитируемой выше статье [28]:

- спонтанная речь (относится к терминополью *речь*),
- социальные факторы (относится к терминополью *социокультурные аспекты*),
- синтаксис (относится к терминополью *грамматика*),
- предложение (относится к терминополью *грамматика*),
- синтагма (относится к двум терминопольям *фонетика, грамматика*),
- фраза (относится к двум терминопольям *речь, фонетика*),
- статистическое моделирование (относится к терминополью *методы исследования*).

На рис. 2 видно, что минимальное количество полей, актуальных для каждой научной статьи исследовательского коллектива 3, а максимальное – 7; мода приходится на 5 полей. Данная модель наглядно показывает значимость исследования *структурной связности* терминопольей научного коллектива, так как структурные связи отражают существующие в реальности комбинации (композиции) научных проблем и направлений.

Графосемантические модели предметной области научного коллектива

Структура терминопольей моделируется с помощью графосемантических моделей. С-граф ПрО, разрабатываемой научным коллективом в 2010 г., представлен на рис. 3, на котором видно, что степень связности вершин графа высока (коэффициент плотности графа 0,463). Ядро графа – терминополье социально-территориальная дифференциация языка (С-ТДЯ), закономерно имеет сильные связи с терминопольем *социокультурные аспекты* (субкультуры, гендер, групповая идентичность, социальный опыт говорящего и многие другие понятия, непосредственно связанные с С-ТДЯ).

В то же время сильная связь терминополья С-ТДЯ (и терминополья *социокультурные аспекты*) с терминопольем *методы исследования* свидетельствует об интенсивной научной рефлексии о методах изучения языка в социолингвистическом и социокультурном измерениях. Значимость остальных терминопольей может быть оценена лишь во временной перспективе; для данного состояния понятийной структуры предметной области оставшиеся терминополья (кроме терминопольей *теория социолингвистики, речь и речь города* также закономерно соотношенных с терминопольями С-ТДЯ и *социокультурные аспекты*) представляют лишь “точки возможного роста”, часть из которых станет заметной в следующих временных срезах.

Графосемантическая модель предметной области, разрабатываемой научным коллективом в 2011 г., представлена на рис. 4, на котором показано, что структура ПрО претерпела значительные изменения. Занимающее прежде ядерное положение терминополье С-ТДЯ сместилось на периферию, а на первый план выдвинулись терминополья *структуры, методы исследования, когниция, социокультурные аспекты* и

ментальный лексикон (одна из актуализовавшихся в 2011 г. “точек роста”, отмеченных на С-графе 2010 г.).

В С-графе на рис. 4 отсутствует явно доминирующее терминополье (как, например, терминополье С-ТДЯ в С-графе 2010 г. – см. рис. 3), более того, сам граф можно рассматривать как становящееся концептуальным пространство с конкурирующими композициями терминопольей. Вероятно, 2011 г. для научного коллектива стал этапом поиска новых перспективных ПрО. Об этом косвенно свидетельствует и уменьшение количества публикаций в рецензируемых журналах (в 1,5 раза меньше по сравнению с 2010 г.), так как новая проблематика требует “вхождения” (осмысления, наработок) в новую / обновленную предметную область. Интересен тот факт, что плотность графа, несмотря на меньшее количество вершин, заметно снизилась по сравнению с 2010 г. – 0,353. Можно предположить, что при достижении определенного порога связности вершин система терминопольей стремится к обновлению, поиску новых ПрО.

Графосемантическая модель предметной области, разрабатываемой научным коллективом в 2012 г., представлена на рис. 5. Разработка научным коллективом новых ПрО осуществлялась достаточно интенсивно: количество публикаций резко выросло (в 2,3 раза) по сравнению с 2011 г. Сама же структура ПрО претерпела заметные изменения.

На рис. 5 видно, что вся структура ПрО состоит из двух относительно самостоятельных подсистем, соединенных друг с другом через ядерное терминополье *социокультурные аспекты*. Интересно, что рефлексия о терминополье *методы исследования* потеряла свою актуальность. Вероятно, это обусловлено разработкой в 2011 г. исследовательских программ изучения новых ПрО. Возникшая в 2012 г. структура ПрО, с одной стороны, сложнее предшествующих структурных состояний, с другой стороны, она неустойчива и, вероятно, подвергнется преобразованию – трансформации малой подструктуры С-графа.

Вероятность такой трансформации обуславливается тем, что данная подструктура в 2012 г. формировалась на основе деятельности только двух участников научного коллектива (рис. 6). Изменение интересов хотя бы одного из этих участников повлечет за собой переупорядочивание этого фрагмента ПрО.

Структурная связность отдельных терминопольей позволяет соотносить способы решения научных проблем. Например, мы уже отмечали интерес к терминополью *лексика* как инструменту изучения терминополья *ментальный лексикон* (на данном этапе такой вывод можно сделать на основании структурных связей между терминопольями на рис. 5).

Картирование научных интересов участников исследовательского коллектива

Структура предметной области научного коллектива в 2010-2012 гг. представлена на рис. 3-5. Однако представляет интерес не только распределение веса терминопольей ПрО по временным срезам, но и выявление “специализаций” участников научного коллектива (в том числе и во временном аспекте).

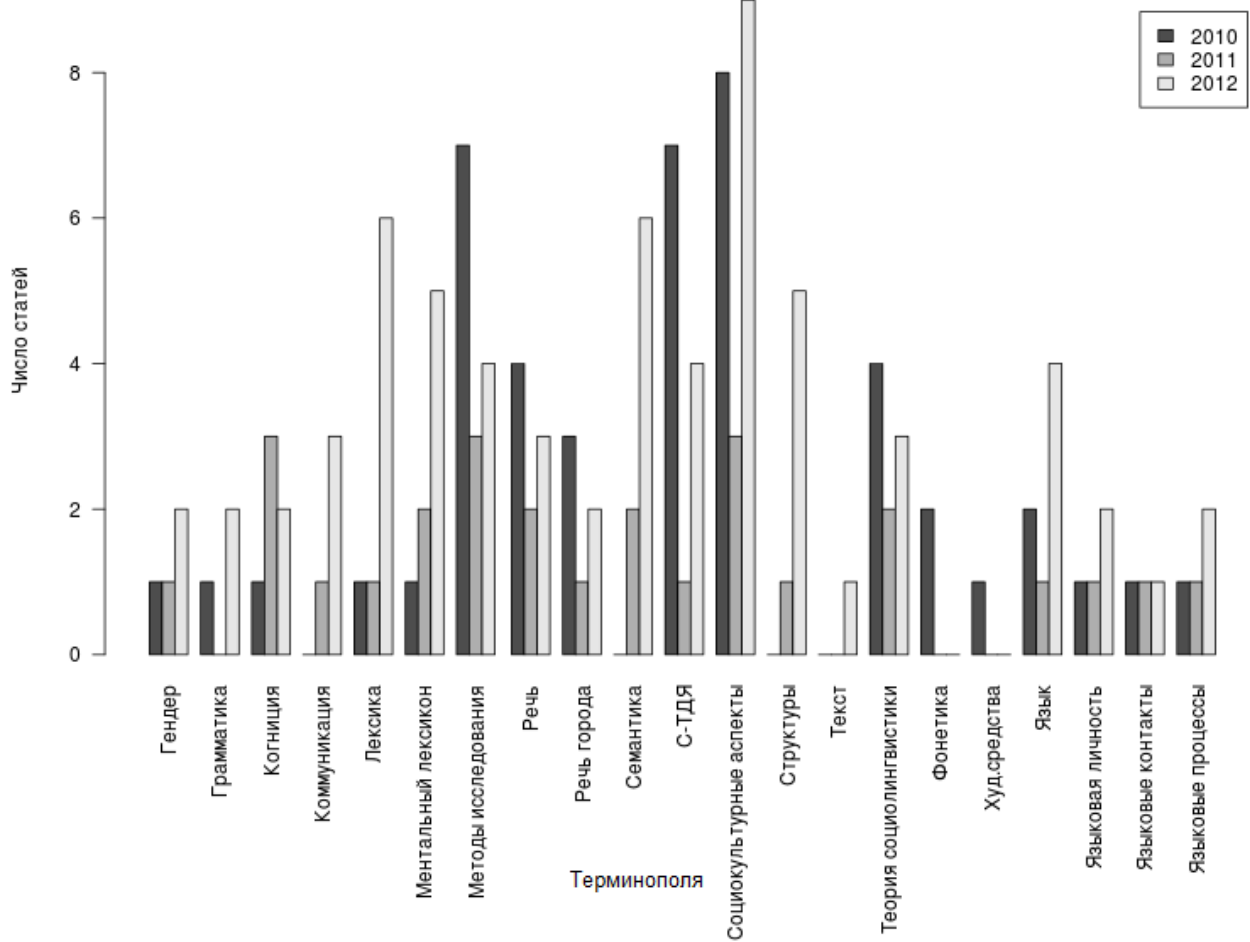


Рис. 1. Распределение частотности терминоплей за период 2010-2012 гг. во всех статьях научного коллектива

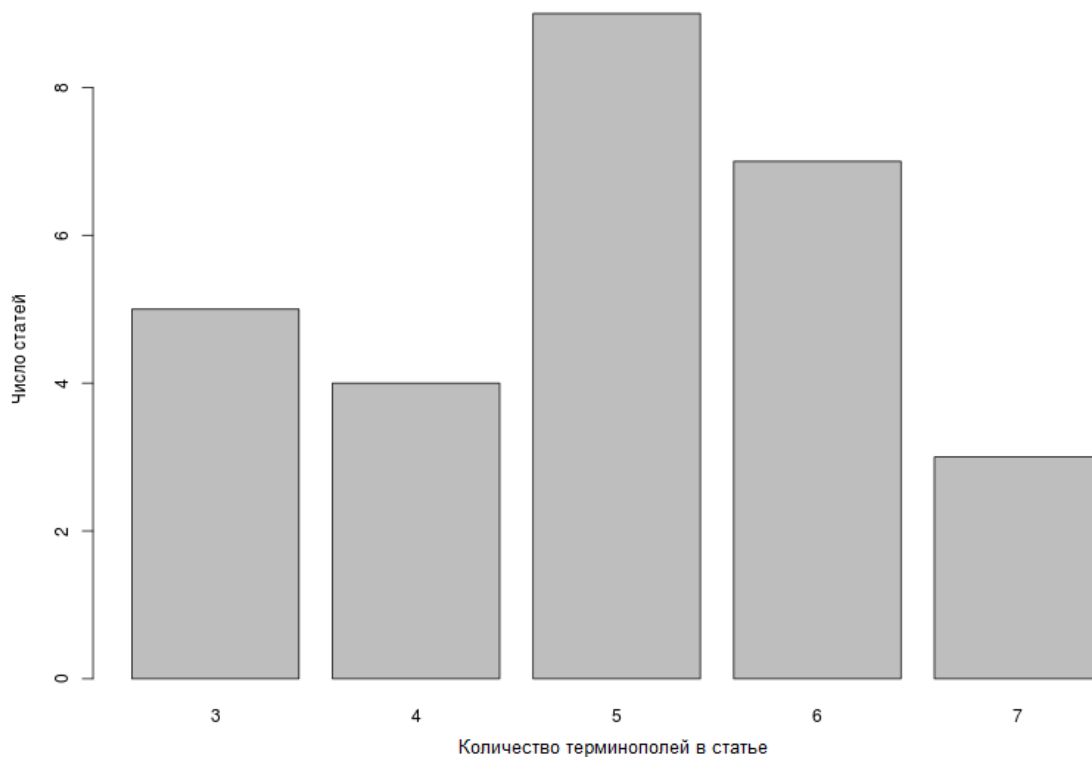


Рис. 2. Распределение количества полей во всех статьях научного коллектива за период 2010-2012 гг.

Примечание. В статье [28] посредством КС актуализировано пять семантических полей (речь, социокультурные аспекты, грамматика, фонетика, методы исследования).

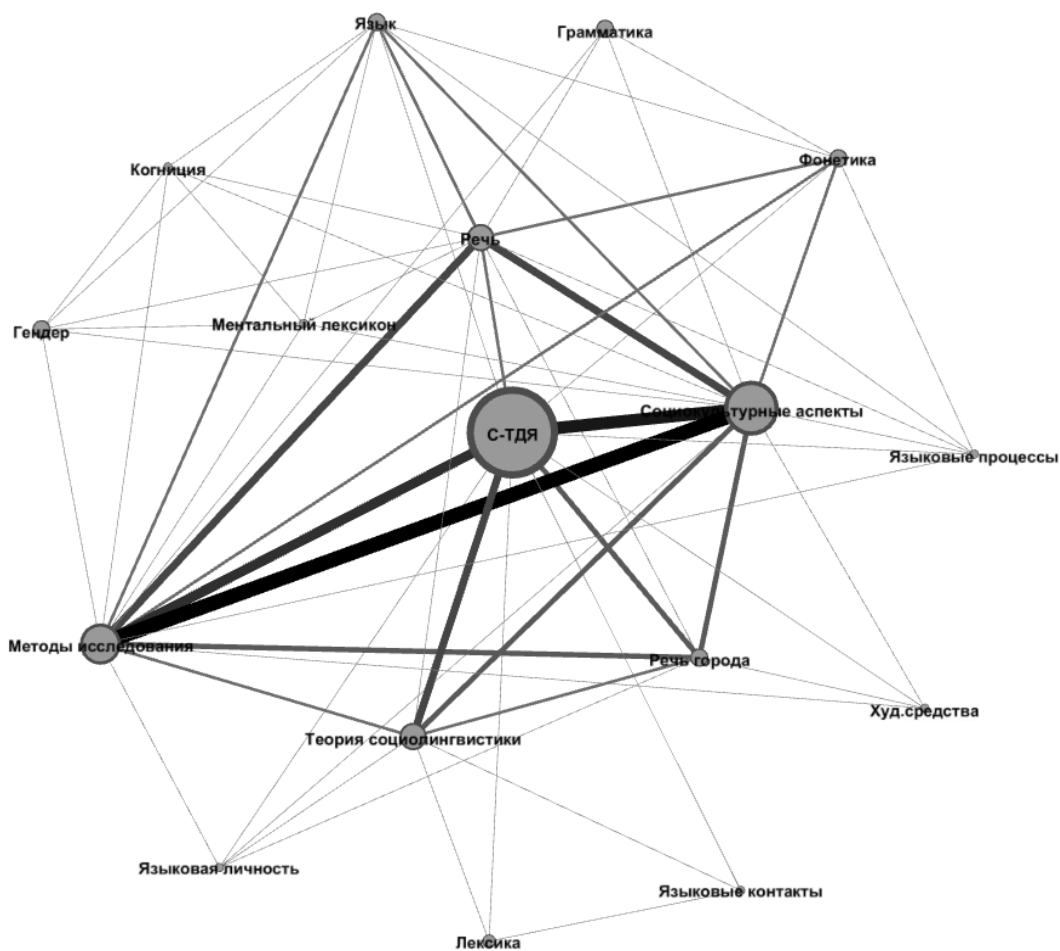


Рис. 3. С-граф предметной области научного коллектива в 2010 г.

Примечание. С-ТДЯ – социально-территориальная дифференциация языка. В С-графе вершины соответствуют терминоплям, а ребра – связям между ними. Размер вершины пропорционален частотности поля, а толщина ребра – силе связи между полями (частоте совместного присутствия двух терминоплей в контекстах).

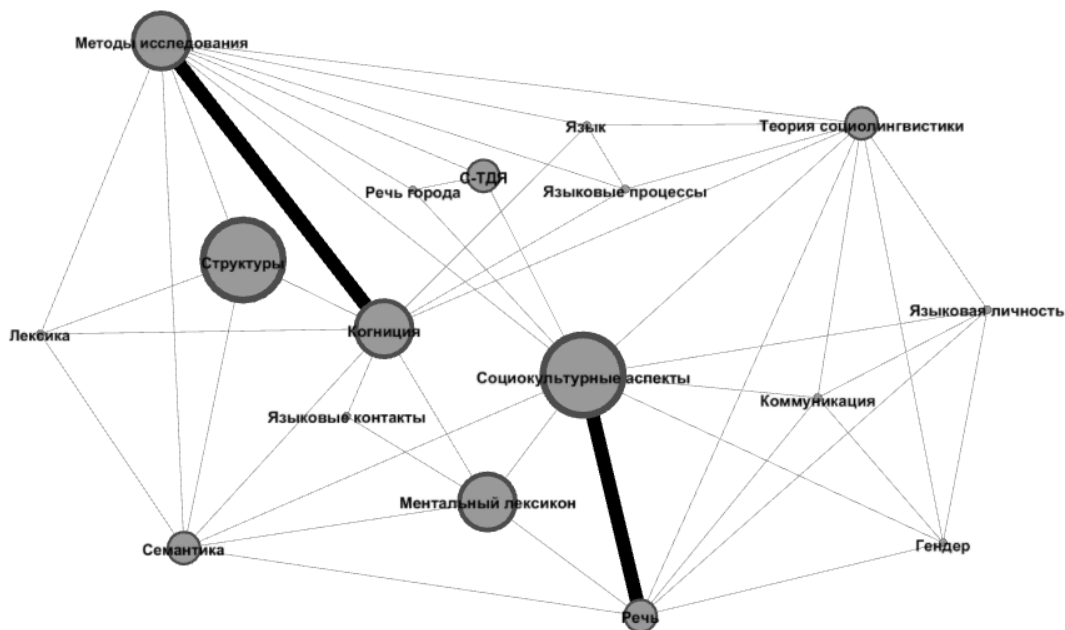


Рис. 4. С-граф предметной области научного коллектива в 2011 г.

Примечание. С-ТДЯ – социально-территориальная дифференциация языка.

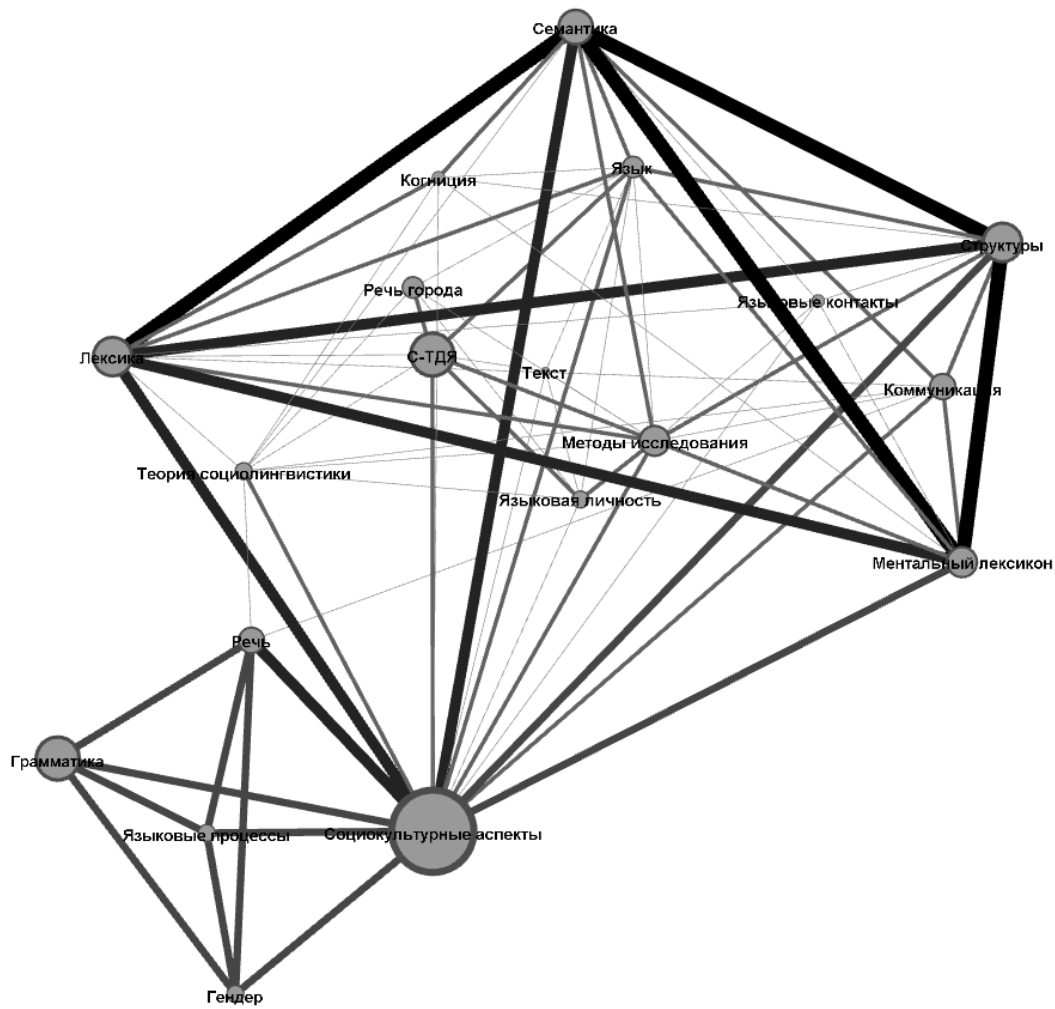


Рис. 5. С-граф предметной области научного коллектива в 2012 г.

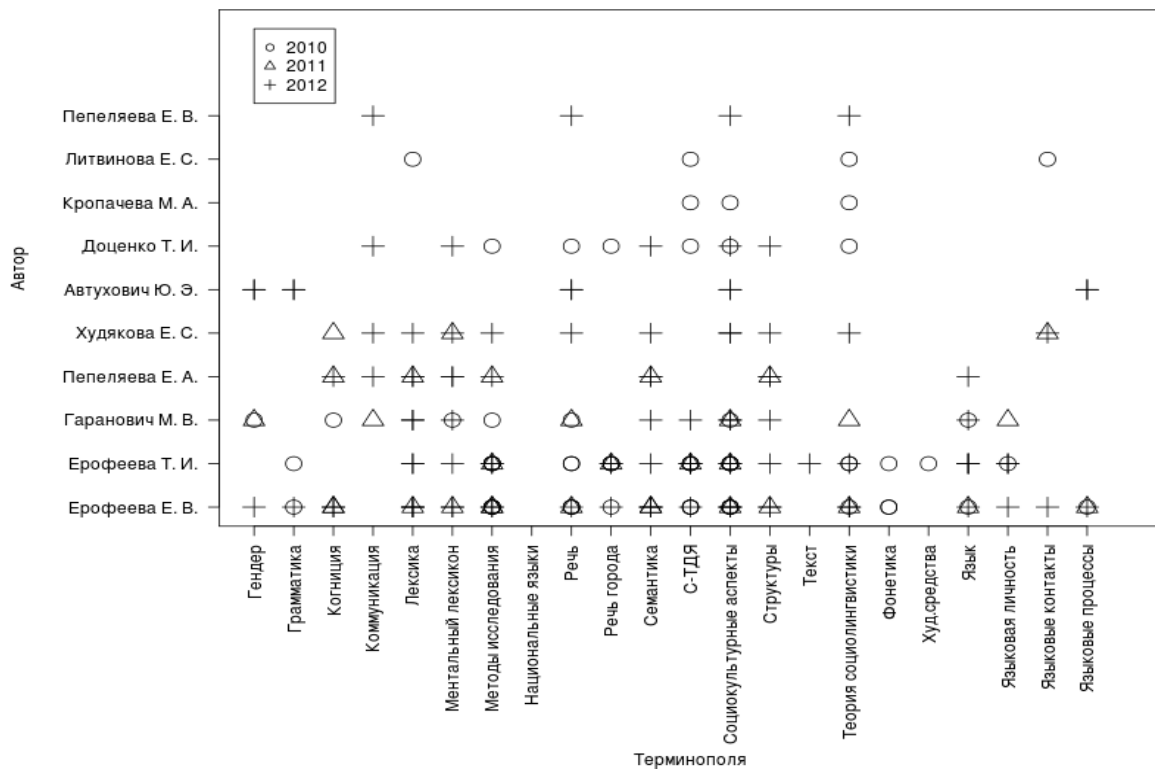


Рис. 6. Карта научных интересов участников исследовательского коллектива в 2010-2012 гг.

Результаты выявленных соответствий между публикациями участников исследовательского коллектива и актуальными для этих публикаций терминоподлями представлены на рис. 6, его можно проинтерпретировать и как карту научных интересов участников исследовательского коллектива, которая дает представление о роли каждого участника научного коллектива в создании интегральной ПрО, о динамике изменения интересов каждого участника и коллектива в целом, о подходах к разработке отдельных направлений в рамках всей ПрО. Карта может быть полезна для планирования совместной деятельности (совместных публикаций, конкурсных заявок и др.).

Частнонаучная предметная область как композиция терминоподлей

Карта научных интересов, представленная на рис. 6, отражает детальное состояние “специализации” участников исследовательского коллектива в разрабатываемой ПрО. Основной ее недостаток состоит в опосредованно данной связности отдельных терминоподлей друг с другом, которая показана через участников исследовательского коллектива, однако, если сферы интересов ученого широки и высока его публикационная активность, то возможность структурного соотнесения отдельных научных проблем друг с другом резко снижается (см. рис. 6). В то же время необходимо построить такую модель, в которой бы отображались научные интересы исследователей в их реальной структурной соотнесенности (в том числе и в аспекте их временной динамики).

Полагаем необходимым более четко обозначить понятие *частнонаучной ПрО*, которую будем понимать как представленную в виде математической модели композицию терминоподлей, репрезентирующую отдельный научный сегмент в общей ПрО исследовательского коллектива (микро-ПрО).

В качестве метода, направленного на выделение частнонаучных предметных областей из общей ПрО научного коллектива, был использован кластерный анализ множества контекстов, параметрами которых являются бинарные векторы, содержащие наборы семантических полей соответствующих контекстов. Набором полей контекста называется множество терминоподлей, связанных с данным контекстом, посредством произвольного числа терминов (в нашем случае ключевых слов к публикациям). В результате кластеризации можно определить: 1) частнонаучные ПрО, соотносимые с выделенными кластерами, и 2) публикации, в той или иной мере соответствующие частнонаучным ПрО.

В нашем исследовании был применен метод нечеткой кластеризации *S-means*. В отличие от других распространенных методов кластерного анализа, таких как *K-means* или самоорганизующихся карт Кохонена, этот метод допускает принадлежность одного элемента двум и более кластерам, что в данной задаче позволяет одну публикацию в случае высокой неопределенности в ее описании отнести к нескольким направлениям. Кроме того, алгоритм метода *S-means* допускает возможность влияния на результат с помощью специального числового параметра

m , $m \in \mathbf{R}$, $m \leq 1$. В качестве исходного значения обычно выбирается $m = 2$, затем в случае неудовлетворительного результата выполняется его подстройка. Итоги этого исследования получены с базовым значением параметра $m = 2$.

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 4-х кластерной модели, которая была отобрана из спектра других моделей (состоящих из 3, 4, 5, 6 и 7 кластеров) на основании наибольшего соответствия ПрО научного коллектива.

Выявленные на основе кластеризации публикаций частнонаучные ПрО представляют собой основные направления исследований научного коллектива в 2010-2012 гг., а представленная на рис. 7 гистограмма распределения выделенных кластеров (частнонаучных ПрО) по временным срезам отражает динамику распределения и перераспределения научных интересов участников исследовательского коллектива.

Несмотря на то, что данные, отображенные на рис. 7, не позволяют осуществлять прогноз исследовательской деятельности (для прогнозирования необходимо иметь временные срезы за более продолжительный период), построенная диаграмма дает возможность оценить, насколько динамична, изменчива и в то же время системна область научных интересов исследовательского коллектива, что проявляется в одновременном воздействии двух факторов: новизны (поиск нового) и традиции. Обновляемость ПрО не означает отказа от уже достигнутого, а скорее подразумевает поиск новых научных контекстов для синтеза разрабатываемых частнонаучных ПрО с уже разработанными частнонаучными ПрО. Например, тот факт, что в 2012 г. резко возросшее количество исследований, связанных с “Социокогнитивными аспектами речевой деятельности” (кластер 1), не привело не только к отказу, но и к снижению доли исследований по другим направлениям, хотя, как мы уже отмечали, 2011 г. был годом поиска нового направления, что отразилось на публикационной активности и, следовательно, на разработанности частнонаучных ПрО. Можно предположить, что новым направлением, конкурирующим с существующими, стало направление “Социокогнитивные аспекты речевой деятельности” (кластер 1).

Поиск новых научных контекстов внешне (на уровне публикационной активности) не сказался на традиционной для исследовательского коллектива частнонаучной ПрО “Социолект и субкультура”. Остальные частнонаучные ПрО демонстрируют единообразную динамику “поведения”: после спада в 2011 г. количества публикаций, в 2012 г. следует либо полное восстановление интенсивности исследований (“Социолект и групповая идентичность” – кластер 2), либо частичное восстановление состояния уровня 2010 г. (“Теория и методология социолингвистических исследований” – кластер 4), что можно объяснить логикой научной деятельности: теоретическое осмысление ПрО осуществляется как “надстройка” над эмпирическим уровнем. Вероятно, что первичных данных и результатов оказалось недоста-

точно для теоретического осмысления новой конфигурации общей ПрО научного коллектива. Однако, оценивая значимость для данного научного коллектива теоретических работ (см. вес кластера 4 в 2010 г. на рис. 7), можно предположить увеличение доли публикаций, посвященных “Теории и методологии социолингвистических исследований” в 2013 г.

Карта научных интересов участников исследовательского коллектива, составленная в соответствии с выделенными частнонаучными ПрО, представлена на рис. 8

В отличие от карты научных интересов, созданной на основе присутствия во всех публикациях научного коллектива тех или иных терминополь (см. рис. 6), карта на рис. 8 отражает: 1) участие каждого исследователя в одном из пяти выделенных направлений и 2) степень близости интересов исследователя тому или иному кластеру (передается в виде расстояния публикации от центра того кластера, к которому она относится).

Карта научных интересов участников исследовательского коллектива во временной динамике (за исследуемый период 2010-2012 гг.) представлена на рис. 9, где видно, как осуществлялся “дрейф” интересов участников научного коллектива, вхождение в новые частнонаучные ПрО. Например, частнонаучная ПрО, оформленная в кластере 1 (“Социокогни-

тивные аспекты речевой деятельности”), становится предметом разработки в основном в 2012 г., а первые работы по ней изданы в 2011 г. При этом, публикация Е.В. Ерофеевой и Е.А. Пепеляевой в 2011 г. [30], относящаяся к данной ПрО, наиболее удалена от центра кластера; в 2012 г. ситуация изменяется, появляются новые работы Е.В. Ерофеевой и Е.С. Худяковой [31], а также Т.И. Доценко [29] и некоторых других – все публикации 2012 г.; они в большей степени соответствуют “образу” ПрО (статьи расположены ближе к центру кластера).

Похожая (но с “противоположным” знаком) ситуация складывается в кластере 4 “Теория и методология социолингвистических исследований”. Доминирующее влияние на “образ” частнонаучной ПрО оказывают работы, созданные в 2010 г.; однако развитие теоретико-методологического базиса социолингвистических исследований требует обращения к новым теоретическим конструктам, в частности, к понятиям из частнонаучной ПрО “Социокогнитивные аспекты речевой деятельности”. Поскольку количество публикаций в 2012 г., относящихся к кластеру 4, уступает совокупности работ по той же частнонаучной ПрО в 2010 г., новые теоретические конструкты пока не играют в ней заметной роли; но при росте публикаций в данной сфере можно ожидать переструктурирования кластера.

Таблица 2

Частнонаучные предметные области (4-х кластерная модель)

№	Кластер	Композиции терминополь	Количество статей	Репрезентативные статьи
1	Социокогнитивные аспекты речевой деятельности	1. Лексика 2. Ментальный лексикон 3. Семантика 4. Социокультурные аспекты 5. Структуры 6. Язык	6	[29, 30, 31]
2	Социолект и групповая идентичность	1. Гендер 2. Грамматика 3. Речь 4. Социокультурные аспекты 5. Языковые процессы	8	[32,33]
3	Социолект и субкультура	1. С-ТДЯ 2. Социокультурные аспекты 3. Теория социолингвистики	6	[34,35]
4	Теория и методология социолингвистических исследований	1. Методы исследования 2. Речь города 3. С-ТДЯ 4. Социокультурные аспекты	8	[36,37, 38]

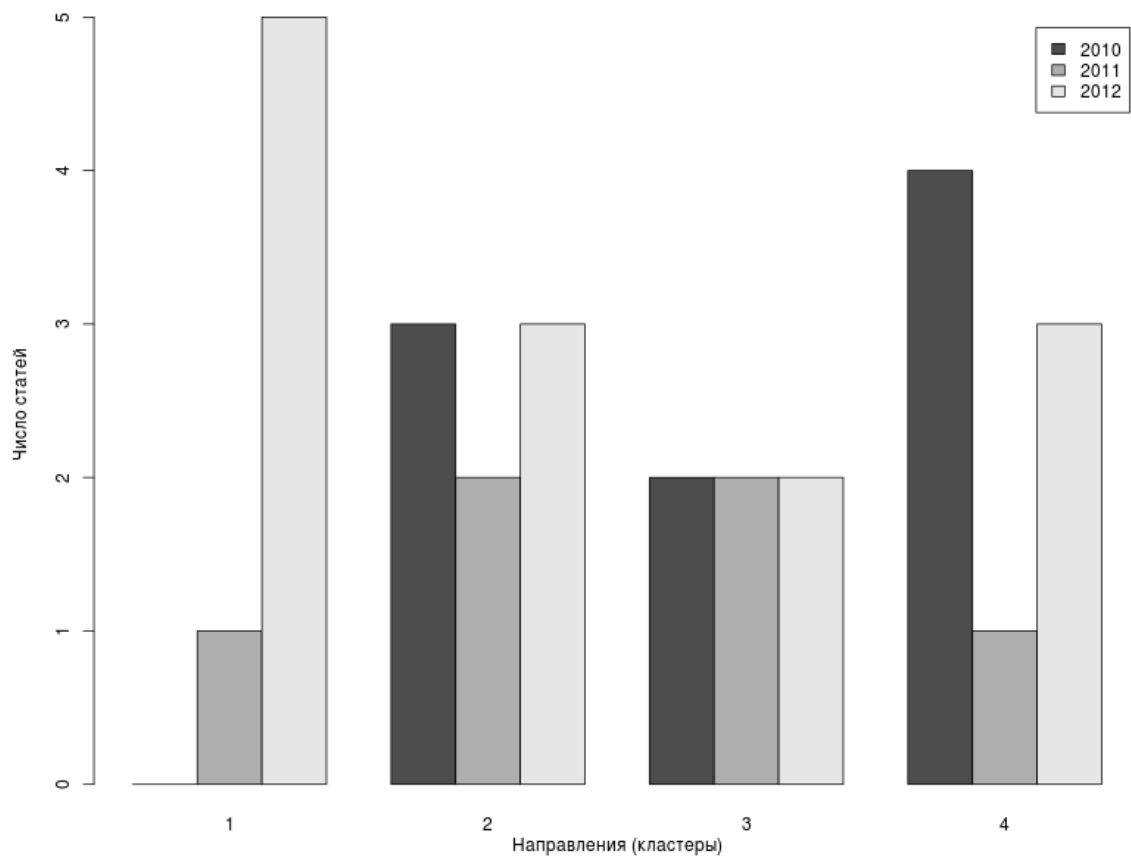


Рис. 7. Распределение выделенных кластеров (частнонаучных предметных областей) во всех статьях научного коллектива по временным срезам 2010, 2011 и 2012 гг.

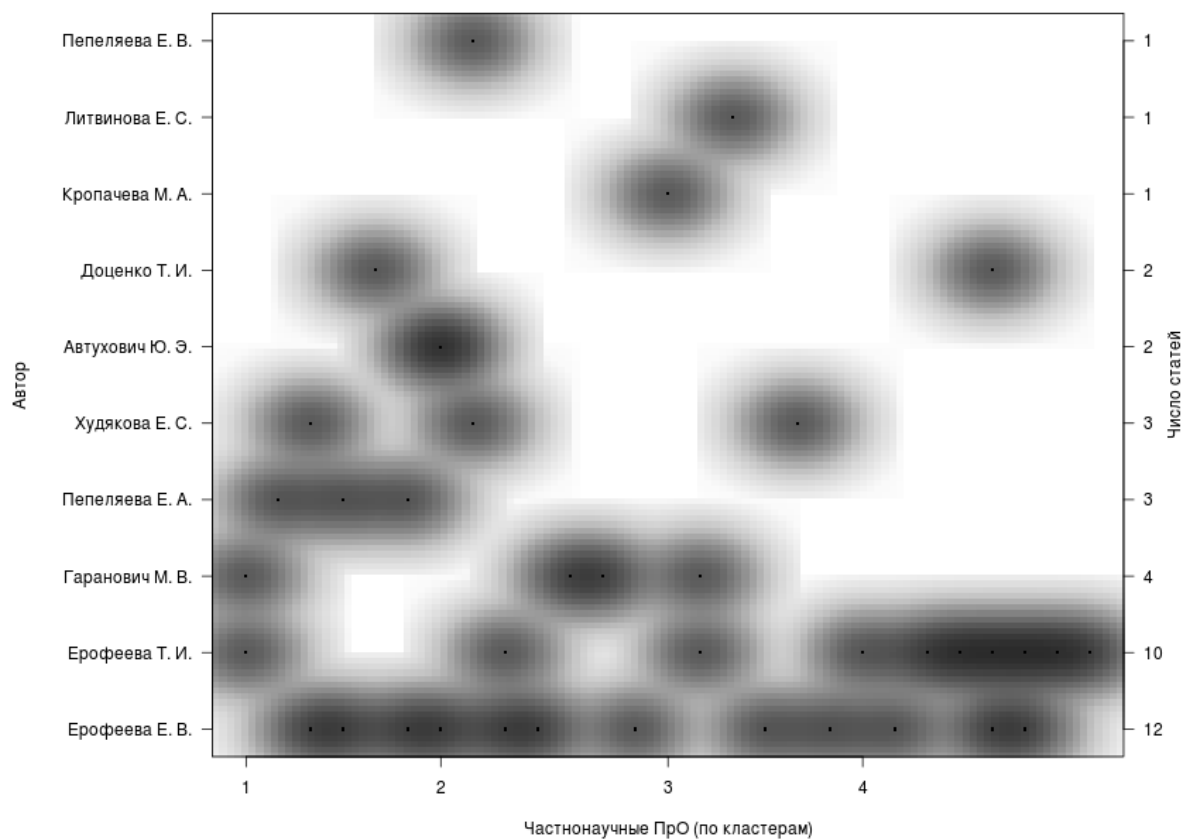


Рис. 8. Карта научных интересов участников исследовательского коллектива (по частнонаучным предметным областям)

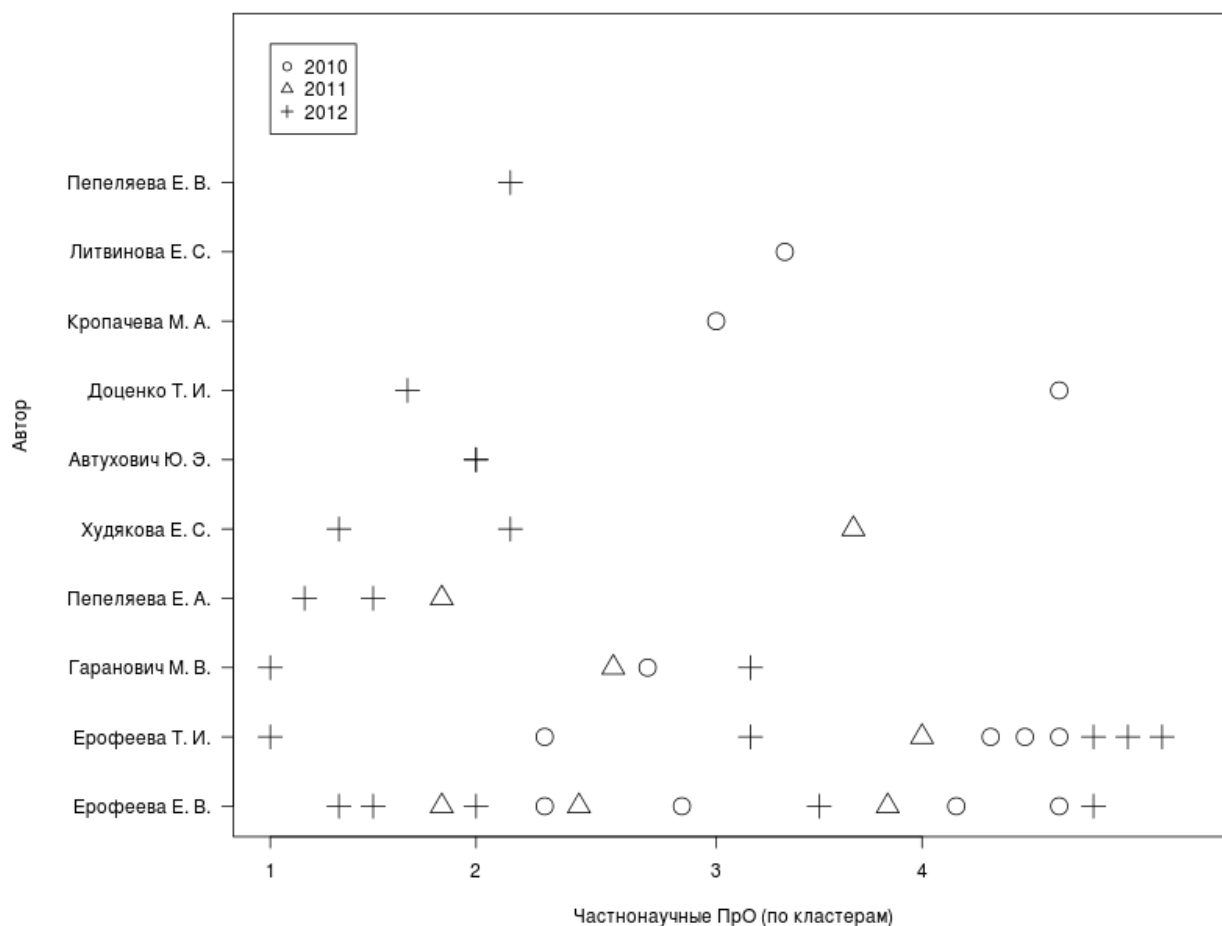


Рис. 9. Карта научных интересов участников исследовательского коллектива (по частнонаучным предметным областям) с учетом временной динамики

ВЫВОДЫ И ПЕРСПЕКТИВЫ

В настоящее время существует множество методов оценки научной деятельности, но отсутствуют доступные инструменты ее эффективного планирования, при котором агент, сообразно своим научным интересам и достижениям, на основании математических моделей развития научной области отбирает частнонаучные ПрО, которые на момент их выбора набирают “вес”. Настоящая статья посвящена описанию первого этапа программы моделирования предметной области научного коллектива на примере деятельности сотрудников лаборатории прикладных и экспериментальных лингвистических исследований Пермской социопсихолингвистической школы за трехлетний период 2010-2012 гг.

Моделирование связной структуры терминопольей осуществлялось с помощью создания графосемантических моделей для каждого временного среза. Выделенные частнонаучные ПрО представляют основные направления исследований научного коллектива, которые могут быть рассмотрены во временной динамике и персональной соотнесенности с каждым сотрудником коллектива. Анализ моделей частнонаучных ПрО позволил сделать вывод о системности научных интересов, проявляющейся в логике развития частнонаучных ПрО относительно общей предметной области научного коллектива.

Полагаем, что моделирование предметной области научного коллектива может быть востребовано для планирования совместной деятельности (совместных публикаций, конкурсных заявок, включения в состав новых участников и др.). Для эффективного планирования научной деятельности необходимо соотносить разрабатываемую научным коллективом ПрО с предметной областью / предметными областями научных направлений, которые разрабатываются всем научным сообществом.

Перспективным в данном аспекте видится разработка методов, инструментов и создание на их основе программных продуктов, позволяющих моделировать структуру частнонаучных ПрО в зарубежных исследованиях, так как, по нашему мнению, среди факторов, препятствующих публикационной активности представителей российской науки (языковые барьеры, ограниченный доступ к БД, агрегирующим научный контент, и др.), центральное место занимает отсутствие определенности в том образе своей предметной области, который сформировался в зарубежной научной среде. Адаптация исследовательского информационного пространства, как отдельного ученого, так и научного коллектива к пространству мировой науки должна включать преодоление неопределенности образа предметной области, в которой работает российский исследователь или научный коллектив.

СПИСОК ЛИТЕРАТУРЫ

1. Egghe L. On the correction of the h-index for career length // *Scientometrics*. – 2013. – Vol. 96, Iss. 2. – P. 563–571.
2. Glanzel W. The role of core documents in bibliometric network analysis and their relation with h-type indices // *Scientometrics*. – 2012. – Vol. 93, Iss. 1. – P. 113–123.
3. Hirsch J.E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship // *Scientometrics*. – 2010. – Vol. 85, Iss. 3. – P. 741–754.
4. 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.
5. Van Leeuwen T.N., Moed H.F., Reedijk J. Critical comments on Institute for Scientific Information impact factors: a sample of inorganic molecular chemistry journals // *Journal of Information Science*. – 1999. – Vol. 25. – P. 489–498.
6. Bordons M., Mauleón E. Male and female involvement in patenting activity in Spain // *Scientometrics*. – 2010. – № 83 – P. 605–621.
7. Costas R., Bordons M. Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective // *Scientometrics*. – 2011. – Vol. 88. – P. 145–161.
8. Mauleón E., Hillán L., Moreno L., Gómez I., Bordons M. Assessing gender balance among journal authors and editorial board members // *Scientometrics*. – 2013. – Vol. 95, Iss. 1. – P. 87–114.
9. Bordons M., Romero F., Zulueta M.A., Barrigón S. Measuring interdisciplinary collaboration within a university: The effects of the multidisciplinary research programme // *Scientometrics*. – 1999. – Vol. 46, Iss. 3. – P. 383–398.
10. Rinia E.J., van Leeuwen Thed N., Bruins Eppo E. W., van Vuren Hendrik G., van Raan Anthony F. J. Measuring knowledge transfer between fields of science // *Scientometrics*. – 2002. – Vol. 54, Iss. 3. – P. 347–362.
11. Van Leeuwen T.N. Strength and weakness of national science systems: A bibliometric analysis through cooperation patterns // *Scientometrics*. – 2009. – Vol. 79, Iss. 2. – P. 389–408.
12. Tijssen R.J.W., Van Leeuwen T.N. Measuring impacts of academic science on industrial research: A citation-based approach // *Scientometrics*. – 2005. – Vol. 66, Iss. 1. – P. 55–69.
13. 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.
14. Коцемир М.Н. Динамика российской и мировой науки сквозь призму международных публикаций // Форсайт. – 2012. – № 1. – С. 38–59.
15. 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.
16. 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.
17. Alekseeva L.M., Isaeva E.V., Mishlanova S.L. Metaphor in Computer Virology Discourse // *World Applied Sciences Journal*. – 2013. – № 27 (4). – P. 533–537.
18. Arano S. Thesauruses and ontologies // *Hipertext.net*. – 2005. – № 3. – URL: <http://eprints.rclis.org/8972/2/12.pdf> (дата обращения: 20.08.2013).
19. 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.
20. Белоусов К.И., Мишланова С.Л., Заседателева М.Г. Тезаурусное моделирование предметной области «Компетентностный подход»: дизайн исследования и программная реализация на платформе ИС «Семограф» // *Инновационные проекты и программы в образовании*. – 2013. – № 4. – С. 3–10.
21. Гладун А.Я., Рогушина Ю.В. Использование онтологических знаний и тезаурусов для объективного профилирования специалистов // *Искусственный интеллект*. – 2006. – № 3. – С. 379–390.
22. Морозова Л.А. Терминознание: Основы и методы. – М.: ГНО «Прометей» МПГУ, 2004. – 144 с.
23. Кузнецов А.М. От компонентного анализа к компонентному синтезу. – М.: Наука, 1986. – 126 с.
24. Щур Г.С. Теории поля в лингвистике. – М.: Эдиториал УРСС, 2009. – 264 с.
25. Система графосемантического моделирования / Д.А. Баранов, К.И. Белоусов, И.В. Влацкая, Н.Л. Зелянская. – М.: Свидетельство о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011.
26. Белоусов К.И., Зелянская Н.Л. Онтология и онтография частнонаучных предметных областей и научная картина мира // *Вестник Пермского университета. Российская и зарубежная филология*. – 2012. – № 4. – С. 104–111.
27. Белоусов К.И., Зелянская Н.Л., Баранов Д.А. Концептуально-гипертекстовая модель управления контентом в ИС «Семограф» // *Вестник Оренбургского государственного университета*. – 2012. – № 11. – С. 59–64.
28. Ерофеева Е.В., Ерофеева Т.И. Человек и текст: антропоцентрический подход к исследова-

- нию // Вестник Пермского университета. Российская и зарубежная филология. – 2010. – № 4. – С. 28-33.
29. Доценко Т.И. Ассоциативно-вербальная сеть в контексте межкультурной коммуникации // Филологические заметки. – 2012. – Ч.1. – С. 147-157.
30. Ерофеева Е.В., Пепеляева Е.А. Структура семантического поля «Человек» в сознании носителей русского языка // Вестник Пермского университета. Российская и зарубежная филология. – 2011. – № 1. – С. 7-19.
31. Ерофеева Е.В., Худякова Е.С. Психолингвистическое исследование ценностных установок билингвов (на материале тематической группы «Человек») // Вестник Пермского университета. Российская и зарубежная филология. – 2012. – № 2. – С. 7-16.
32. Ерофеева Е.В. Дифференциация и интеграция социолектов: к вопросу о взаимодействии социальных факторов // Вестник Пермского университета. Российская и зарубежная филология. – 2010. – № 5. – С. 39-47.
33. Ерофеева Е.В. Групповая идентичность и ее отражение во внутреннем лексиконе // Вестник Санкт-Петербургского университета. Серия 9: Филология. Востоковедение. Журналистика. – 2011. – № 1. – С. 91-96.
34. Ерофеева Е.В., Автухович Ю.Э. Влияние гендерного фактора на выбор видовых форм глагола // Глобальный научный потенциал. – 2012. – № 18. – С. 143-145.
35. Гаранович М.В., Ерофеева Т.И. Словари городских субкультурных образований // Глобальный научный потенциал. – 2012. – № 18. – С. 139-142.
36. Ерофеева Е.В. Влияние специальности на структуру языковой картины мира // Филологические заметки. – 2012. – Ч.2. – С. 229-245.
37. Доценко Т.И., Ерофеева Е.В., Ерофеева Т.И. Пермская школа социолингвистики: теоретические и методологические основания // Вестник Пермского университета. Российская и зарубежная филология. – 2010. – № 2. – С. 144-155.
38. Ерофеева Т.И. Локальная единица в городской литературной речи как результат диалога двух языковых систем // Вестник Ленинградского государственного университета им. А.С. Пушкина. – 2010. – Т.1. – С. 71-78.
39. Ерофеева Т.И. Социолект как инструмент описания языковой ситуации региона // Вестник Пермского университета. Российская и зарубежная филология. – 2010. – №1. – С. 21-25.

Материал поступил в редакцию 17.01.14.

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

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