

ТЕМАТИЧЕСКАЯ И ПАРАДИГМАЛЬНАЯ МОДЕЛИ КОНЦЕПТОСФЕРЫ НАУКИ (на материале перспективных исследований русской лингвистики)*

Белоусов Константин Игоревич – доктор филологических наук, профессор. Пермский государственный национальный исследовательский университет. Российская Федерация, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: belousovk@gmail.com

Баранов Дмитрий Александрович – научный сотрудник. Пермский государственный национальный исследовательский университет. Российская Федерация, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: baranov@semograph.com

Ерофеева Елена Валентиновна – доктор филологических наук, профессор. Пермский государственный национальный исследовательский университет. Российская Федерация, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: elenerofee@gmail.com

В статье представлено описание двух подходов к моделированию концептосферы науки: тематического и парадигмального. В качестве материала исследования был взят корпус рефератов инициативных проектов, поддержанных отделением гуманитарных и общественных наук РФФИ в области языкознания, а также корпус рефератов статей исследователей (и их соавторов), получавших неоднократную поддержку научного фонда. Тематическое моделирование осуществлялось в рамках двух подходов: от корпуса текстов (моделирование концептосферы как целого в виде сети терминопольей) и от единичного текста (выделение композиций терминопольей, устойчиво реализуемых в научных проектах, которые могут рассматриваться как отдельные направления языкознания). Базовыми методами исследования послужили метод графосемантического моделирования, реализованный в Информационной системе «Семограф» (подход «от корпуса»), и метод кластеризации K-means (подход «от текста»). В качестве операционных единиц выступили сгруппированные в терминополья наборы ключевых слов, отраженные в рефератах проектов, поддержанных научным фондом. Парадигмальное моделирование базировалось на анализе сетевого взаимодействия исследователей (проведенном на основе информации о совместных публикациях). После создания гиперграфа исследователей (2108 вершин) с помощью метода модулярности (аналога кластерного анализа для сетевых моделей) осуществлялось разбиение графа на подграфы – сетевые сообщества. Каждый кластер можно рассматривать как модель научного сообщества, реализующего определенную научную парадигму, а совокупность кластеров – как модель концептосферы предметной области с точки зрения представленности и взаимосвязи отдельных парадигм. Перспективой моделирования концептосферы науки является синтез тематической и парадигмальной моделей. Представленные модели описания концептосферы науки могут использоваться как одно из средств



* Статья подготовлена при поддержке РФФИ, проект № 15-06-06373 «Когнитивно-информационные модели исследовательской деятельности научных коллективов».



мониторинга, прогнозирования и управления научными исследованиями, т. е. как инструмент государственной и/или ведомственной политики.

Ключевые слова: концептосфера науки, термины, терминополь, тематическая модель, парадигмальная модель, семантический граф, кластеризация

THEMATIC AND PARADIGM MODELS OF THE CONCEPT SYSTEM OF SCIENCE (a case study of advanced research in Russian linguistics)

Konstantin I. Belousov – DSc in Philology, professor.
Perm State University.
15 Bukirev St., Perm, 614990,
Russian Federation;
e-mail: belousovki@gmail.com

Dmitriy A. Baranov – research fellow.
Perm State University.
15 Bukirev St., Perm, 614990,
Russian Federation;
e-mail: baranov@semograph.com

Elena V. Erofeeva – DSc in Philology, professor.
Perm State University.
15 Bukirev St., Perm, 614990,
Russian Federation;
e-mail: elenerofee@gmail.com

The article describes two approaches to modeling the concept system of science – the thematic and paradigm ones. The research represents a case study of the two corpora of abstracts: abstracts of projects supported by the Department of Humanities and Social Sciences of the Russian Federal Property Fund in linguistics, as well abstracts of articles by authors (and their co-authors) who have received multiple support from this foundation. Thematic modeling was carried out within the frameworks of two approaches: corpus based approach (modeling the system of concepts as a holistic entity of term fields network), and single text approach (singling out compositions of term fields steadily present in the texts of projects which can be treated as separate branches of linguistics). The method of semantic graph modeling realized in the “Semograph” Information system (corpus based approach) and the K-means clustering method (single text based approach) were applied. Arrays of keywords (those that occurred in the abstracts of projects supported by the Foundation) grouped into term fields served as operational units. Paradigmatic modeling was based on the analysis of network interaction of researchers (created on the basis of the information on joint publications). After the hypergraph of researchers (2,108 state points) had been created, it was divided into subgraphs (network communities) by means of the modularity method (analogue of cluster analysis). Each cluster can be considered as a model of a scientific community that actualize a certain scientific paradigm; a set of clusters represents a model of the concept system of a certain subject domain from the point of view of representation and interrelation of several paradigms. The synthesis of thematic and paradigmatic models appears to be the direction for future research that involves modeling the concept system of science. The considered models can be applied as an effective means of monitoring, forecasting and managing scientific research, i.e. as an instrument of state and/or department policy.

Keywords: concepts system of science, terms, fields of terms, thematic model, paradigm model, semantic graph, clustering



Введение. Качественный и количественный анализ лингвистической парадигмы

Научную предметную область мы понимаем как 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 предметной области: конкуренция (изменение конфигураций, зависимостей, влияния друг на друга) частнонаучных предметных областей в рамках общей дисциплинарной предметной области, а также обновление концептуального пространства предметной области за счет междисциплинарных связей.

Внутренние границы науки являются более подвижными, чем внешние, что проявляется, с одной стороны, в представлениях о предметах научной дисциплины, сформированных в ее отдельных областях (например, в языкознании речь идет о предметах структурной лингвистики, психолингвистики, социолингвистики, лингвокультурологии и мн. др.). Но с другой стороны (и на наш взгляд, более важной), ***внутренние границы науки проводятся вероятностно всякий раз в каждом отдельном исследовании.*** Иными словами, внутренние границы науки есть вероятностный феномен, доступный изучению на основе анализа корпуса/корпусов научных публикаций, рефератов научных проектов, патентов и др.

Сложность научной предметной области, возникающей за счет а) **возрастающего объема конкурирующих концепций, теорий, направлений, методов и др.**, б) усиливающейся междисциплинарности и в) происходящей глобализации науки (размывании национальных исследовательских парадигм особенно в гуманитаристике), – приводит к тому, что с помощью одного качественного анализа сложно, а может быть и невозможно, релевантно описать и объяснить процессы, происходящие в науке.

В лингвистике, как, вероятно, и в любой другой науке, анализ прошлого и современного состояний исследовательских парадигм является предметом постоянной рефлексии [Кибрик, 1995; Кубрякова, 1995]. Однако основным подходом к структурированию языковедческих направлений остается качественный анализ, основанный на наиболее востребованных работах в области общего, частного и прикладного языкознания и собственном исследовательском опыте, эрудиции, интересах и предпочтениях ученых.



На наш взгляд, главным недостатком качественного подхода (не смотря на его детальность и глубину) является даже не субъективная трактовка организации предметной области науки, а дискретное представление информационного пространства научной дисциплины. Анализ ограниченного количества работ не дает возможности показать континуум коллективной научной деятельности и порождает представления о сосуществовании в современной теоретической лингвистике антагонистических теорий языка [Кошелев, 2013]. Не ставя под сомнение мысль автора, заметим, что в научной реальности кажущиеся противопоставленными концепции могут сложным образом взаимодействовать в отдельных научных публикациях или даже в публикациях отдельной научной школы.

Именно поэтому представляет интерес не только качественный, но и количественный анализ взаимодействия научных идей, концепций, направлений. Под количественным анализом в данном случае понимается **моделирование концептуального пространства выбранной предметной области**, а не наукометрические методы, основанные либо на показателях цитирования, либо на экспертных оценках. Формальные наукометрические параметры в настоящее время подвергаются обоснованной критике: утверждается, что показатели цитируемости исследователей, импакт-факторов журналов часто не соответствуют истинному положению дел ни в отношении самой статистики, ни в отношении качества исследований и порождают негативные процессы в науке, такие как взаимное цитирование, множество форм манипулирования импакт-фактором (редакционное давление на авторов, «взращивание» собственных авторов, написание обзорных статей с многочисленными ссылками и др.), дискриминацию национальных наук, агрессивное гендерное поведение и др. [Adler et al., 2011; Арнольд, Фаулер, 2011; Варшавский и др., 2011; Дыбо, Крылов, 2013; Lawrence, 2006; Sugimoto, Cronin, 2013, и др.]. Кроме того, показатели цитируемости и многочисленные инструменты, созданные на их основе (индексы Hirsch h-index, Hirsch h-core, g-index, R-index; AR-index, Jin A-index и др.) не дают никакой содержательной информации о процессах, идущих в научной предметной области, и ее прошлых, настоящих и будущих структурных состояниях. Экспертный анализ (метод экспертных оценок) [Арнольд, Фаулер, 2011; Ding, Cronin, 2011], безусловно, интересен, но крайне затратен, так как связан с необходимостью привлечения многочисленных общепризнанных экспертов в различных областях научного знания.

Целями количественного анализа парадигмы лингвистических исследований становятся:

1) **выявление основных направлений лингвистических исследований на основе анализа корпусов научных текстов;**



- 2) установление связей частнонаучных направлений, репрезентированных в отдельных научных текстах;
- 3) прогнозирование состояний выделенных научных направлений и определение наиболее востребованных из них;
- 4) рассмотрение выделенных научных направлений с точки зрения их «взгляда» на предмет науки, методы, цели и результат научного исследования.

Научная предметная область как концептосфера

Предметную область мы рассматриваем как взаимосвязанную совокупность концептов, т. е. как концептосферу [Лихачев, 1993, р. 161]. Структурными единицами концептосферы научной предметной области являются концепты, формируемые на основе научных понятий, выражаемых терминами. Единицы терминологического пространства предметной области благодаря разным контекстам употребления концептуализируются, «обрастают» новыми значениями, не отражаемыми в процессе кодификации.

Концептуализация термина осуществляется в процессе понимания и овладения/присвоения термина и стоящего за ним понятия концептуальной системой индивида. Процесс понимания (как «присвоение чужого») можно рассматривать следующим образом: «Усвоить некоторый смысл (концепт) – значит построить некоторую структуру, состоящую из имеющихся концептов в качестве интерпретаторов, или анализаторов, рассматриваемого концепта, “вводимого” – с внешней точки зрения, т. е. с точки зрения некоторого наблюдателя, находящегося вне системы, – в... конструируемую систему концептов, или концептуальную систему» [Павиленис, 1983, с. 100]. Очевидно, что в концептосфере науки можно говорить о разных типах концептов: парадигмальных, аксиологических, деятельностных, поведенческих, оказывающих влияние как на конкретный исследовательский процесс, так и на развитие науки в целом.

Предлагаемая в данной работе модель предметной области не является ее терминологической онтологией [Biemann, 2005] – в нашем случае речь идет одновременно и о кодифицированных, и о некодифицированных единицах (которых значительно больше, чем кодифицированных), зависящих от ситуации и контекста. Термины, используемые в научных публикациях, представляют собой отдельные лексемы или словосочетания, часть которых нельзя отнести ни к терминосистеме, ни даже к терминологии изучаемой предметной области. Одна (причем довольно объемная) часть терминов представляет собой общенаучный слой терминологии (**категория, система, понятие** и др.),



другая содержит специальную лексику, характерную для других наук или сфер деятельности (**российская блогосфера, телепроект, автомобилестроение, компетентностный подход** и др.). И наконец, внушительная по объему часть терминов относится не к терминологии, а к номенклатуре (**лезгинский язык, названия дворянских усадеб** и др.). Однако термины и другие наименования, не относящиеся напрямую к исследуемой предметной области, крайне значимы для нее: без этого предметно-понятийного субстрата невозможно ни представить синхронный срез предметной области, ни предложить ее ретроспективную или перспективную модель развития.

Мы исходим из того, что концептосфера научной предметной области представляет собой сложную, открытую, динамичную систему, и ее анализ может осуществляться с опорой на разные подходы к моделированию. На наш взгляд, можно говорить, по меньшей мере, о двух моделях описания концептосферы: тематической и парадигмальной.

Тематическая модель концептосферы науки

Тематическая модель описания концептосферы основана на выделении терминопoley и установлении отношений между ними.

Терминопole определяется как множество терминов, объединенных общей семантикой [Реформатский, 1959; Суперанская, и др., 2012]. Отказ от анализа собственно гиперсети терминов обусловлен слабыми интерпретационными возможностями, которые дают аналоги методов кластеризации данных, например, метод модулярности графов [Blondel et al., 2008; Lambiotte, Delvenne, Barahona, 2009], позволяющий выделить кластеры (в терминах метода модулярности – сообщества) в гиперсети терминов. Вероятно, слабый интерпретационный потенциал выделенных кластеров обусловлен обсуждаемой выше концептуализацией терминов – закреплением за ними новых смыслов благодаря новым контекстам использования понятий. Таким образом, логика исследования приводит к необходимости экспертной классификации эмпирически наблюдаемых данных (терминов). Результатом такой классификации являются терминопole.

Терминопole – явление понятийной сферы, представляющее собой категорию экстралингвистической действительности; их структура, содержание и границы обусловлены логикой тех фрагментов действительности, которые они описывают, а также логикой научного исследования. Терминопole может относиться ко всей предметной области или ее части. С помощью него упорядочивается и хранится разнородная информация о предмете, материале, методах



исследования и др. Репрезентацией терминопольей на языковом уровне могут быть не только термины данной предметной области, но и термины других предметных областей разного объема: «терминополье – это арена активного функционирования и взаимодействия терминов из разных терминосистем, в том числе и из других областей знаний. И между такими терминами (и терминологическими подсистемами) возможны образные сближения и заимствования» [Виноградов, 2014, с. 368]. **Терминополья предметной области могут многократно пересекаться, образуя единую сложную динамическую систему, способную перестраиваться в зависимости от внешних параметров.** Таким образом они создают сетевые структуры более высокого уровня, чем сетевые структуры эмпирически наблюдаемых единиц – отдельных терминов предметной области (подробнее см. [Белоусов, Ерофеева, Зелянская, 2017; Belousov, Baranov, Zelyanskaya, 2014]).

В качестве материала для тематического моделирования концептосферы науки в данном случае используются корпуса научных текстов или их реферативных комплексов¹.

Тематическое моделирование концептосферы предметной области может осуществляться в рамках следующих подходов:

Подход «от корпуса текстов»

Данный подход основан на надстройке над гиперсетью терминов, используемых в научных работах, **графа терминопольей**.

Связи между терминопольями устанавливаются на основе совместного присутствия в одном контексте (например, наборе ключевых слов к статье) двух и более терминов, относящихся к разным терминопольям. Выявленные связи между всеми терминопольями во всем корпусе данных (или отдельной выборки внутри корпуса) можно представить в виде семантической карты – матрицы $N \times N$, в которой в столбцах и строках располагаются названия терминопольей. Значение ячейки, находящейся на пересечении столбца и строки, отражает частоту совместной встречаемости двух терминопольей в контекстах корпуса (или его выборки). На основе семантической карты строится семантический граф, который представляет собой визуализацию семантической карты. Узлами семантического графа являются терминополья, ребрами – связи между терминопольями. Объемы терминопольей и частоты связей графически передаются размером (для узла) и толщиной (для ребер). Семантический граф рассматривается как модель предметной области.

Подход «от корпуса» в качестве дополнительных возможностей предполагает использование разнообразных фильтров метаданных, описывающих научные публикации (страна/город, научный статус,

¹ В данном случае в качестве текста используются реферат статьи, состоящий из заголовка, аннотации и ключевых слов.



цитирование, журнал и др.). В данном случае появляется возможность генерации семантического графа, представляющего какой-либо аспект (например, определенный город и/или журнал и т. п.) в сети терминопольей концептосферы предметной области.

На рис. 1 представлен семантический граф терминопольей, репрезентирующий перспективные исследования в отечественной лингвистике (материалом послужили поддержанные РФФИ научные проекты в области лингвистики в 2015 г.). Видно, что граф имеет сложную структуру с отсутствием резких границ между ядром и периферией; практически отсутствуют терминополья с маленькими валентностями (количеством связей с другими полями). Для предметной области традиционно значимы терминополья, называющие единицы анализа и материал исследования (**лексикология, грамматика, семантика, языки мира**), однако наблюдается стремление обозначить направление (**когнитивная лингвистика, ареальная лингвистика, социолингвистика** и др.) и методологию исследований (**методы**). Обращает на себя внимание значительное количество проектов, представляющих современные научные направления, возникшие на стыке лингвистики с другими науками (**социолингвистика, психолингвистика, этнолингвистика, лингвокультуроло-**

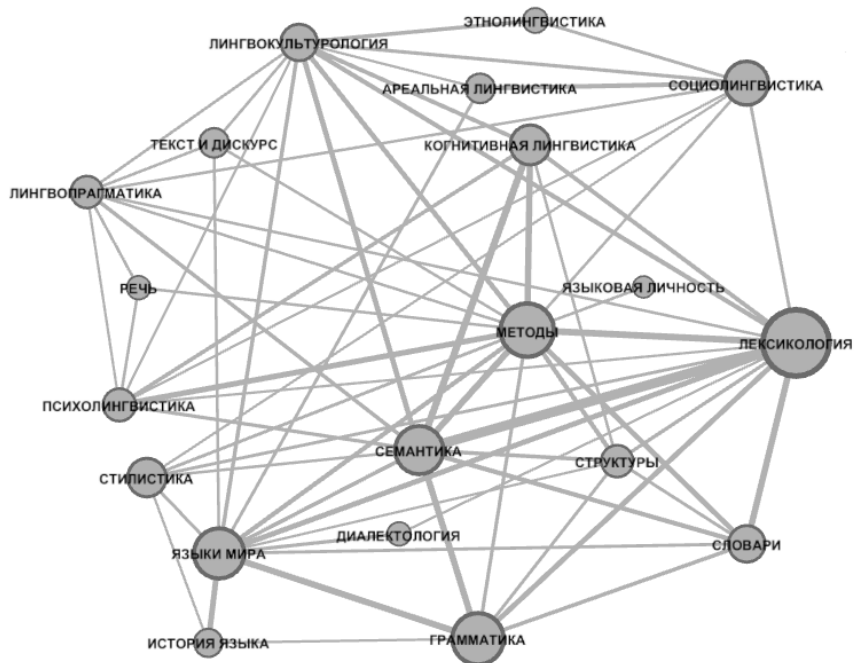


Рис. 1. Граф терминопольей предметной области ЯЗЫКОЗНАНИЕ 2015 г.



гия, лингвопрагматика и др.). Заметно влияние на предметную область когнитивных и психолингвистических исследований, активное обращение к проблемам теории коммуникации и прагматики, а также текста и дискурса. Кроме того, следует выделить формирование устойчивого запроса на информационные продукты, представляемые по результатам НИР (**словари, базы данных, корпусы**).

Таким образом, семантический граф терминопольей 2015 г. демонстрирует сбалансированное распределение весов терминопольей, т. е. наличие конкуренции между научными направлениями. Граф терминопольей является вероятностным графом, состояние которого обусловлено предшествующими временными состояниями, что позволяет прогнозировать состояния графа терминопольей в краткосрочной и/или долгосрочной перспективе.

Подход «от текста к корпусу»

Описанный выше метод структурирования предметной области с помощью графосемантического моделирования основывается на парной связности всех терминопольей друг с другом, однако набор терминов в конкретных исследованиях (публикациях) может относиться более чем к двум терминопольям. Это делает актуальным разработку такого метода структурирования предметной области, который бы учитывал композиции терминопольей каждой отдельной публикации и выявлял структуру всех таких композиций в корпусе текстов.

Для этого вводится понятие *частнонаучной предметной области* – представленной в виде математической модели композиции терминопольей, репрезентирующей отдельный научный сегмент в общей предметной области [Belousov, Baranov, Zelyanskaya, 2014]. В качестве метода выделения частнонаучных предметных областей используется кластерный анализ (метод K-means). Кластерный анализ производится на множестве контекстов (в нашем случае, наборов ключевых слов), каждый из которых представлен как бинарный вектор $C_i = (P_{1i}, P_{ni})$, где i – номер контекста, P_{ji} – значение (0 или 1 в зависимости от отсутствия/присутствия в i -м контексте) j -го терминополья. В результате кластеризации определяются: 1) частнонаучные предметные области, соотносимые с выделенными кластерами, и 2) проекты, в той или иной мере соответствующие частнонаучным предметным областям.

Кластерный анализ позволяет разделить множество контекстов корпуса на несколько групп (кластеров). С помощью метода кластерного анализа можно варьировать количество разбиений корпуса, т. е. получать спектр n -кластерных моделей, из которого выбирается репрезентативная модель частнонаучных направлений предметной области. Репрезентативность модели определяется рядом параметров: наибольшим охватом терминопольей, представленностью наиболее



частотных терминополь, отсутствием дублирующих друг друга кластеров и т. п. Кластерная модель репрезентируется в виде таблицы с перечислением терминополь, входящих в каждый кластер, обозначением количества публикаций, относящихся к каждому кластеру. Модель может учитывать временную динамику, пространственную локализацию и др. параметры, описывающие публикации (см. рис. 2).

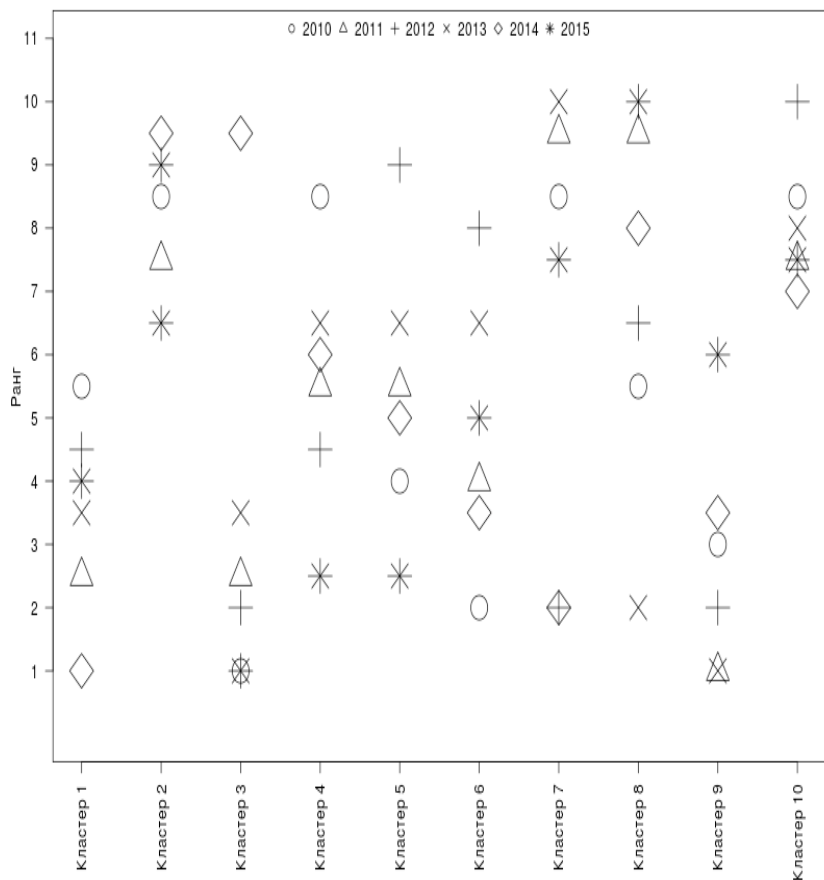


Рис. 2. Распределение выделенных кластеров (частнонаучных предметных областей) по временным срезам 2010–2015 гг.

Примечание. Ранги даны в порядке убывания значимости (количества проектов)

Кластер 1 – **ареальная лингвистика и языковые контакты + социолингвистика**

Кластер 2 – **ареальная лингвистика и языковые контакты + диалектология + лексика и лексикология + словари, базы данных, корпусы + языки мира**



Кластер 3 – лексика и лексикология + методы + семантика + словари, базы данных, корпусы

Кластер 4 – стилистика, поэтика, риторика + материал исследования

Кластер 5 – когнитивная лингвистика + лингвокультурология + семантика + языки мира

Кластер 6 – грамматика + языки мира

Кластер 7 – когнитивная лингвистика + психолингвистика + речь + семантика + текст и дискурс + прагматика

Кластер 8 – грамматика + лексика и лексикология + фонетика + языки мира

Кластер 9 – история языка + лексика и лексикология + языки мира

Кластер 10 – диалектология + лексика и лексикология + лингвокультурология + этнолингвистика

На рис. 2. показано временное распределение поддержанных проектов, относящихся к выделенным частнонаучным предметным областям. Так как общее количество поддержанных проектов заметно варьируется во времени, количественные показатели были ранжированы (чем выше частота данного направления, тем ниже ранг). Таким образом, чем меньше сумма всех рангов для того или иного направления, тем выше его поддержка Фондом. Еще одним показателем является величина коридора колебаний (отрезок между минимальным и максимальным рангами для каждого кластера), показывающая разброс между долями кластеров поддержанных проектов РФФИ в разные годы.

Наибольшая поддержка Фонда в области инициативных исследовательских проектов в целом оказывается направлению 9 – «**история языка + лексика и лексикология + языки мира**», направлению 3 – «**лексика и лексикология + методы + семантика + словари, базы данных, корпусы**» и направлению 1 – «**ареальная лингвистика и языковые контакты + социолингвистика**». Данные направления включают как традиционные для лингвистики области (**история языка, словари**), так и относительно новые междисциплинарные сферы (**базы данных, корпусы, языковые контакты, социолингвистика**). Но, несмотря на то что отдельные направления поддерживаются несколько лучше других, в целом не прослеживается четких временных тенденций в предпочтении поддержки каких-либо частнонаучных областей в ущерб другим. Например, направление 3, по которому в 2010 и 2015 гг. было поддержано наибольшее количество проектов, в 2014 г. такой поддержки не получило.

В целом тематическое моделирование концептосферы науки в рамках двух обозначенных подходов позволяет а) выделять актуальные для определенного временного среда терминополья, б) определять доминирующие научные направления, в) прогнозировать состояние концептосферы науки. Инструмент прогнозирования может исполь-



зоваться как средство управления научными исследованиями, т. е. как инструмент государственной и/или ведомственной политики (например, большая поддержка междисциплинарных проектов определенной тематики).

Парадигмальная модель концептосферы науки

Тематическая модель концептосферы науки в единстве двух подходов позволяет выделять основные научные направления, представленные как композиции тех или иных терминопольей. Однако данная модель «уравнивает» способы **видения** предметов исследования и варианты решения научных проблем разными учеными – т. е. **парадигмальные подходы**². Между тем исследователи могут использовать одни и те же термины, но понимать их по-разному; кроме того, сами подходы к проведению исследований в одной и той же области могут заметно различаться. В то же время ученый может опираться на один парадигмальный подход при смене не только узкой темы, но и направления. Например, количественный подход актуален для когнитивистики, лингвокультурологии, социолингвистики и т. д. – от исследователя требуется только наличие соответствующих знаний.

Выделение актуальных исследовательских парадигм, вероятно, наиболее просто осуществить, если «привязать» каждую парадигму к группе ученых, ведущих совместную научную деятельность: количественный анализ публикаций показывает, что в настоящее время 90 % всех научных и инженерных публикаций принадлежат двум и более авторам, большинство статей написано от 6 до 10 авторами из более чем одного учреждения [Cooke, Hilton, 2015].

Предполагается, если два и более исследователя работают вместе над одной научной проблемой, то они имеют сходные точки зрения на предмет, методы и результат исследования, что отражается в совместных публикациях по итогам работы. Таким образом, проблеме выделения актуальных исследовательских парадигм легко свести к проблеме выделения групп ученых, ведущих совместную научную деятельность. На наш взгляд, решение данной проблемы можно осуществлять двумя способами.

² Говоря о научной парадигме, мы исходим из ее понимания Т. Куном, полагая при этом, что данное понятие уместно использовать не только для диахронического сопоставления научных эпох, но и для синхронического анализа состояния науки.



Научные коллективы

Если понимать группы исследователей как относительно устойчивые **научные коллективы**, то для определения границ и состава их участников следует обращаться либо к экспертным суждениям членов научных коллективов, либо использовать формальный признак аффилиации ученых [Adams et al., 2005]. Однако обращение к понятию научного коллектива для анализа концептосферы всей научной предметной области имеет много недостатков, возникающих вследствие сложности определения границ научных коллективов, которые могут заметно меняться даже на протяжении небольшого временного отрезка. При этом обращение к суждениям участников одного и того же научного коллектива может дать разное понимание границ коллектива даже в синхронии. На размытость границ научных коллективов влияет и то, что в современном научном пространстве исследователь может входить сразу в несколько научных групп, т. е. взаимоотношения ученого и научной группы могут быть разные. Ситуация с неопределенностью границ научных коллективов усугубляется и тем, что научные группы зачастую создаются для реализации конкретных проектов. Такие группы имеют заранее определенные временные рамки и не ограничены территориально (т. е. участники научного проекта могут работать удаленно).

Сказанное позволяет сделать вывод, что обращение к экспертным суждениям о границах научного коллектива уместно только в случае исследования конкретных коллективов; использование же формального признака места работы исследователей для определения границ научного коллектива подменяет само понятие научного коллектива на вторичные (с точки зрения науки) понятия кафедры, факультета, института и др., которые могут быть востребованы только в целях анализа эффективности/успешности научной деятельности данных административных единиц, что обычно и является целью наукометрических исследований [Hirsch, 2010; Tijssen, van Leeuwen, 2005 et al.].

Моделирование научных сообществ

Если понимать группы исследователей как относительно устойчивые сообщества, возникающие вследствие близости понимания предмета, цели, методов исследования и научного результата отдельными учеными, то для определения границ и состава участников сообществ можно обратиться к анализу сетевого взаимодействия исследователей, проводимому на основе информации о совместных публикациях.



Мы исходим из того, что соавторство является формой проявления *научной идентичности*, которую понимаем как разделяемую субъектом систему взглядов, относящихся а) к науке в целом и к предметной области, в которой специализируется исследователь (в данном случае речь идет как о эпистемологических, так и об аксиологических аспектах); б) к пониманию сущности научного исследования (включая представления о существующих концепциях, о постановке научной проблемы, об исследовательском материале, о первичных данных, о системе методов, о научном результате и др.); в) к проведению исследования и представлению его результатов. Можно сказать, что научная идентичность характеризуется через полипарадигмальную отнесенность исследователя. Мы говорим о полипарадигмальности, т. к. в действительности любое исследование включено одновременно в множество парадигм. Например, антропоцентрическая парадигма, в соответствии со стандартами которой может выполняться исследование, не отвергает возможности квантитативного подхода к проведению исследования; в рамках того же антропоцентризма могут существовать и такие парадигмы, как когнитивная, коммуникативно-дискурсивная и др. В этом смысле современная лингвистика (как и вся гуманитаристика) представляется набором парадигм разного объема, находящихся во всех существующих типах семантических отношений. Такое многообразие парадигм (как глобальных, типа системно-структурной или антропоцентрической, так и локальных – например, парадигм, создаваемых научными школами) затрудняет процесс формирования у субъекта научной идентичности, поскольку многие представления, разделяемые исследователем, могут не осознаваться как парадигмальные.

В то же время совместная исследовательская деятельность (особенно в случае междисциплинарных работ) приводит к выработке *парадигмального консенсуса*, т. е. разделяемой двумя и более учеными системе/совокупности мнений о предмете, методах, результатах исследования. Парадигмальный консенсус в данном случае отражает идею такого пересечения парадигм, которое приводит к появлению относительно устойчивой и определенной в своих границах и составе *локальной парадигмы*. При этом тот факт, что исследователь А осуществляет совместную научную деятельность с исследователем Б свидетельствует об исходной близости их парадигмальных установок. В этом смысле для автора любой соавтор является «инструментом» формирования научной идентичности.

Таким образом, вместо обращения к понятию научного коллектива с a priori данным списком научных коллективов и их участников, мы приходим к возможности моделирования исследовательских групп на основе построения гиперграфа исследователей, созданного на материале совместных публикаций. Узлами такого гиперграфа являются ученые, а ребрами отображается наличие совместных науч-



ных работ. Сетевой анализ позволит выделить исследовательские сообщества, каждое из которых основано на собственном парадигмальном консенсусе и имеет свою исследовательскую парадигму.

В нашей работе в качестве первичных узлов графа рассматривались исследователи-лингвисты, чьи инициативные проекты были неоднократно поддержаны отделением гуманитарных и общественных наук РФФИ. Остальные узлы графа достраивались на основе совместного авторства публикаций. Глубина сети имеет значение 3, т. е. поиск соавторов проводился по схеме Автор – Соавтор – Соавтор – Соавтор. В результате общее количество узлов графа (исследователей) составило 2108; общее количество совместных публикаций – более 45 тыс. При анализе не учитывались коллективные монографии и сборники с большим числом соавторов. Количество соавторов в одной работе – не более пяти.

После создания гиперграфа исследователей с помощью метода модулярности графа (аналога кластерного анализа) осуществлялось разбиение графа на подграфы – сетевые сообщества (см. рис. 3). Выделенные группы исследователей интерпретировались как парадигмально отличающиеся друг от друга. В данном случае концептосфера предметной области членилась на основе близости парадигмальной, а не тематической.

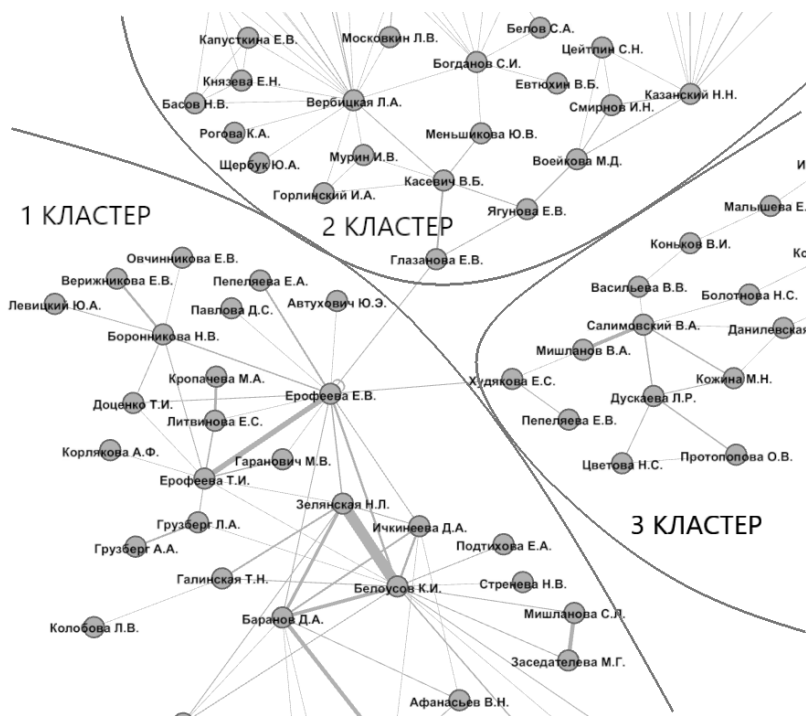


Рис. 3. Участок гиперграфа исследователей (фрагменты с тремя кластерами)

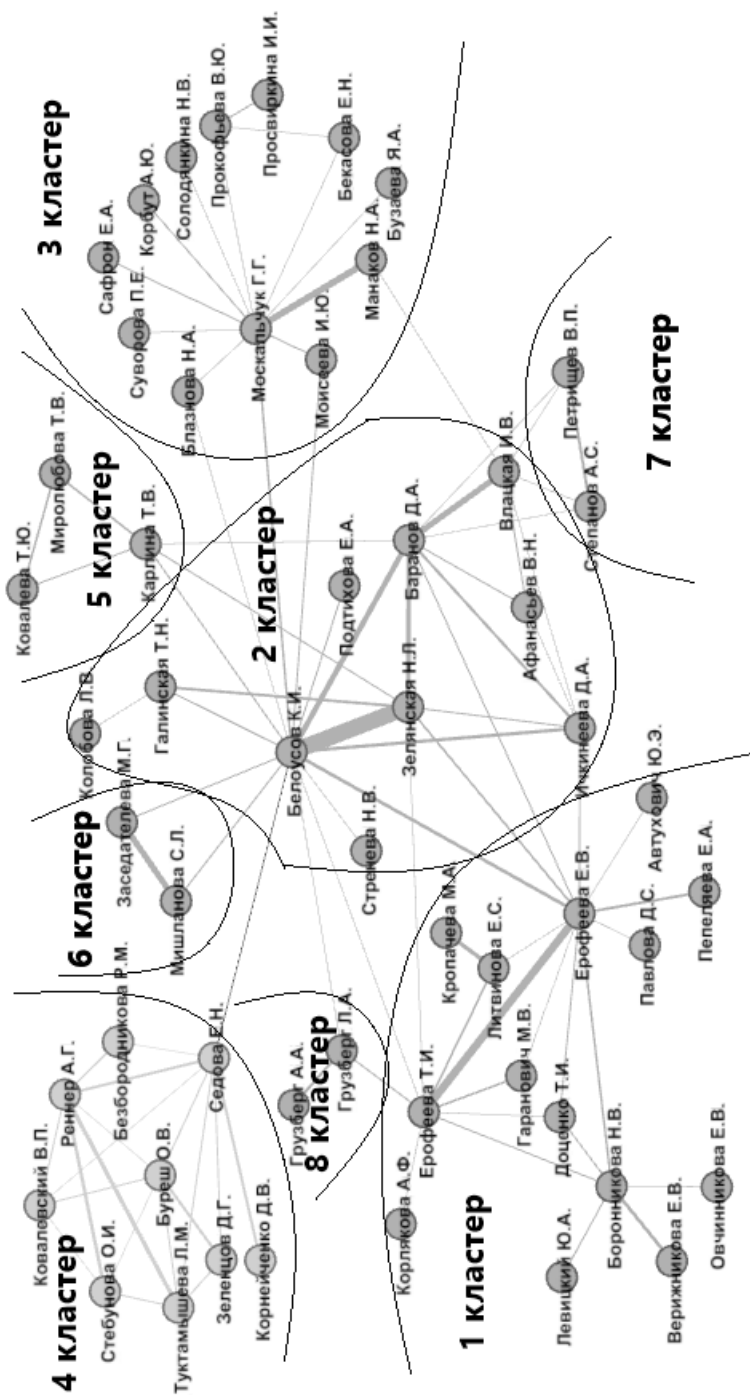


Рис. 4. Результат применения операции модулярности к Кластеру 1 (рис. 3)



Выявленные на уровне всего гиперграфа кластеры в дальнейшем могут быть снова пройти кластеризацию (см. рис. 4). В этом случае мы получим вероятностные модели сообществ, каждая из которых реализует собственную локальную *частнонаучную парадигму*.

На рис. 4 видно, что Кластер 1 (рис. 3) распадается на 8 кластеров разного объема: от двух до 14 исследователей. Часть кластеров состоят только из представителей лингвистических наук (кластеры 1, 3, 6, 8). В кластер 2 входят как лингвисты, как и специалисты в области прикладной математики и статистики; кластеры 4 и 5 представляют исследователи-экономисты; а кластер 7 – ученые-географы. Включенность в определенные кластеры ученых – представителей других наук позволяет выявить междисциплинарный потенциал предметной области.

Каждый конкретный кластер можно рассматривать как модель научного сообщества, реализующего определенную научную парадигму, а совокупность кластеров как модель концептосферы предметной области с точки зрения представленности и взаимосвязи отдельных парадигм.

Перспективы исследования

Перспективой моделирования концептосферы науки является синтез тематической и парадигмальной моделей.

Кластеризация авторов позволяет закрепить за каждым кластером список публикаций (в нашем случае – выборки реферативных комплексов). Используя созданную на первом этапе (в рамках тематического анализа) классификацию терминов по терминоплям, можно создавать парадигмальные модели терминоплей и направлений для каждого кластера (локальной парадигмы).

При моделировании концептосферы науки также представляется перспективным обращение к иным способам классификации терминологии. Например, особый интерес вызывает возможность идентификации терминов как монодисциплинарных, междисциплинарных или общенаучных.

Помимо собственно терминов при парадигмальном подходе особую значимость приобретает анализ библиографического списка научной работы: выбора источников, количества цитируемых работ, их жанров, глубины цитирования статей, языка источников, количества самоцитирований и т. п. Сама природа научного текста имеет «гипертекстовый» характер, поэтому анализ присутствия чужого слова в тексте позволяет моделировать парадигмальную преемственность.

Особым аспектом, представляющим интерес, является практико-ориентированность исследования, которую мы интерпретируем как возможность использования результатов работы в различных отраслях



экономики. Отнесение публикации к той или иной рубрике экономических классификаторов определяет статус исследования как прикладного. Можно рассматривать данный вопрос с двух точек зрения. С одной стороны, можно получить статистику вклада определенной предметной области (или парадигмального кластера внутри предметной области) в разные сектора экономики. С другой стороны, можно получить статистику вклада разных предметных областей (или парадигмальных кластеров внутри предметной области) в один из секторов экономики.

Список литературы

- Белоусов, Ерофеева, Зелянская, 2017 – *Белоусов К.И., Ерофеева Е.В., Зелянская Н.Л.* О внешних и внутренних границах современной лингвистики (графосемантическое моделирование концептосферы предметной области на основе корпусного подхода) // *Вопр. когнитивной лингвистики*. 2017. № 4. С. 45–57.
- Виноградов, 2014 – *Виноградов В.А.* Термин в научном дискурсе // *Вестн. Нижегород. ун-та им. Н.И. Лобачевского*. 2014. № 2 (1). С. 368–372.
- Кибрик, 1995 – *Кибрик А.Е.* Современная лингвистика: откуда и куда // *Вестн. МГУ. Сер. Филология*. 1995. № 5. С. 93–103.
- Кошелев, 2013 – *Кошелев А.Д.* Современная теоретическая лингвистика как Вавилонская башня (о «мирном» сосуществовании множества несовместимых теорий языка) // *Изв. РАН. Сер. лит. и яз.* 2013. Т. 72. № 6. С. 3–22.
- Кубрякова, 1995 – *Кубрякова Е.С.* Эволюция лингвистических идей во второй половине XX в. (опыт парадигмального анализа) // *Язык и наука конца 20 века. М.: Ин-т языкознания РАН*, 1995. С. 144–238.
- Лихачев, 1993 – *Лихачев Д.С.* Концептосфера русского языка // *Известия РАН. Сер. лит. и яз. М.: Ин-т языкознания РАН*, 1993. Т. 52. № 1. С. 3–9.
- Павиленис, 1983 – *Павиленис Р.И.* Проблема смысла: современный логико-философский анализ языка. М.: Мысль, 1983. 286 с.
- Реформатский, 1959 – *Реформатский А.А.* Терминология. Введение в языкознание. М.: Учпедгиз, 1959. 432 с.
- Суперанская, Подольская, Васильева, 2012 – *Суперанская А.В., Подольская Н.В., Васильева Н.В.* Общая терминология: вопросы теории. М.: ЛИБРОКОМ, 2012. 248 с.
- Adams et al., 2005 – *Adams J.D., Black G.C., Clemmons J.R., Stephan P.E.* Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999 // *Research Policy*. 2005. Vol. 34. Iss. 3. P. 259–285. URL: <http://www.sciencedirect.com/science/article/pii/S0048733305000132> (дата обращения: 07.05.2016).
- Belousov, Baranov, Zelyanskaya, 2014 – *Belousov K.I., Baranov D.A., Zelyanskaya N.L.* A research team and its subject area: Towards the question of the effective planning of scientific activities // *Scientific and Technical Information Processing*. 2014. Vol. 41. Iss. 2. P. 85–97.
- Biemann, 2005 – *Biemann C.* Ontology Learning from Text: A Survey of Methods // *Journal for Language Technology and Computational Linguistics*. 2005. Vol. 20. P. 75–93.



Blondel et al., 2008 – *Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E.* Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment*. 2008. Vol. 10. URL: <http://iopscience.iop.org/article/10.1088/17425468/2008/10/P10008/pdf> (дата обращения: 12.06.2016).

Cooke, Hilton, 2015 – Enhancing the Effectiveness of Team Science // *Committee on the Science of Team Science; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Research Council* / Ed. by Cooke N.J., Hilton M.L. Washington DC: The national Academies Press, 2015. 256 p.

Hirsch, 2010 – *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.

Lambiotte, Delvenne, Barahona 2009 – *Lambiotte R., Delvenne J.-C., Barahona M.* Laplacian Dynamics and Multiscale Modular Structure in Networks. 2009. URL: <https://arxiv.org/abs/0812.1770v3> (дата обращения: 16.05.2016).

Tijssen, van Leeuwen et al., 2005 – *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.

References:

Adams, J. D., Black, G. C., Clemmons, J. R., Stephan, P. E. “Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999”, in: *Research Policy*, 2005, vol. 34, iss. 3, pp. 259–285. [<http://www.sciencedirect.com/science/article/pii/S0048733305000132>, accessed on 07.05.2016]

Belousov, K. I., Baranov, D. A., Zelyanskaya, N. L. “A research team and its subject area: Towards the question of the effective planning of scientific activities”, *Scientific and Technical Information Processing*, 2014, vol. 41, iss. 2, pp. 85–97.

Belousov, K. I., Erofeeva, E. V., Zelyanskaya, N. L. “O vneshnih i vnutrennih granichah sovremennoj lingvistiki (grafosemanticheskoe modelirovanie konceptosfery predmetnoj oblasti na osnove korpusnogo podhoda)” [External and internal boundaries of contemporary linguistics (corpus based semantic graph modeling of concepts sphere of subject domains)], *Voprosy Kognitivnoy Lingvistiki*, 2017, no. 4, pp. 45–57. (In Russian)

Biemann, C. “Ontology Learning from Text: A Survey of Methods”, *Journal for Language Technology and Computational Linguistics*, 2005, vol. 20, pp. 75–93.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, 2008, vol. 10. [<http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/pdf>, accessed on 12.06.2016]

Cooke, N. J., Hilton, M. L. (eds.). “Enhancing the Effectiveness of Team Science”, in: *Committee on the Science of Team Science; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Research Council*. Washington DC: The national Academies Press, 2015. 256 pp.



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, pp. 741–754.

Lambiotte, R., Delvenne, J.-C., Barahona, M. *Laplacian Dynamics and Multiscale Modular Structure in Networks*. 2009. [<https://arxiv.org/abs/0812.1770v3>, accessed on 16.05.2016].

Likhachev, D. S. “Konceptosfera russkogo yazyka”, *Izvestiya RAN. Seriya literatury i yazyka*, 1993, vol. 52, no. 1, pp. 3–9. (In Russian)

Pavilenis, R. I. *Problema smysla: sovremennyy logiko-filosofskiy analiz yazyka* [Problem of meaning: modern logical-philosophical analysis of language]. Moscow: Mysl’, 1983. 286 pp. (In Russian)

Reformatskiy, A.A. *Terminologiya. Vvedenie v yazykoznanie* [Terminology. Introduction to linguistics]. Moscow: Uchpedgiz, 1959. 432 pp. (In Russian)

Superanskaya, A. V., Podol’skaya, N. V., Vasil’eva, N. V. *Obshchaya terminologiya: Voprosy teorii* [General terminology: questions of theory]. Moscow: LIBROKOM, 2012. 248 pp. (In Russian)

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, pp. 55–69.

Vinogradov, V.A. “Termin v nauchnom diskurse” [A term in scientific discourse], *Bulletin of Lobachevsky University of Nizhni Novgorod*, 2014, no. 2 (1), pp. 368–372. (In Russian)