

ЯЗЫКОЗНАНИЕ

И.Д. ЧЕСНОКОВА, М.Е. МАНЬШИН
(Волгоград)

НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА КАК ОСНОВНОЙ ИНСТРУМЕНТ ПОИСКА ПРИ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ (на примере поиска антонимов в публицистических текстах)

Описаны возможности осуществления различных лингвистических исследований средствами Национального корпуса русского языка. В качестве примера поиска необходимых единиц, оптимизированного по времени и объему охвата текстового материала, представляется алгоритм поиска пар антонимов в публицистических текстах.

~~~~~

Ключевые слова: *лингвистическое исследование, корпус, материал исследования, Национальный корпус русского языка, антонимы, алгоритм поиска.*

Лингвистические исследования независимо от поставленных задач ориентированы на применение широких массивов языковых употреблений. Чем шире количественный охват речевого проявления того или иного языкового явления, тем выше точность результатов проведенного исследования и достовернее проявление наблюдаемых закономерностей и законов.

Традиционно сбор материала исследования связан с процессом выборки исследуемых единиц из лексикографических источников и текстов различных стилей и жанров, а также с опросом информантов по разным методикам, проведением полевых записей и т. д. Одновременно с этим исследователь проводит ручную обработку текстов и занесение их на физический носитель, чаще всего им являются карточки, которые, в свою очередь, составляют картотеки, существующие на бумаге либо в форме электронной среды. В случае фикса-

ции звучащей речи используются технические средства.

Поиск и обработка информации в больших картотеках при традиционном подходе (тем более если перед исследователем стоят задачи обновления материала или изучения динамики того или иного явления) требуют значительного количества времени. В связи с вышесказанным, а также с закрытостью личных карточек для научного сообщества и потребностью в большом текстовом материале многие ученые, занимающиеся прикладными аспектами лингвистической науки, видят перспективу в широком использовании различных компьютерных сред, баз данных и поисковых систем в методике лингвистического исследования (Ю.А. Авдеев, А.Б. Кутузов, В.В. Мамонтова, Е.В. Падучева и др.) [1; 12; 14; 19].

При проведении исследований, которые своим предметом имеют лексическую систему языка, ускорение процесса получения языкового материала, использование сети Интернет в качестве дистанционного доступа, увеличение объема информации, быстрый доступ к ней и ее обработка в полной мере позволяют задействовать преимущества корпусов. Это связано с особенностями лексической системы как подсистемы языка: наиболее открытая и способная к изменениям, она реализует его эпистемическую функцию. Оптимизацией данной функции наряду с лексикографией и терминологией занимается корпусная лингвистика [4, с. 112]. Являясь частью прикладной лингвистики, корпусная лингвистика основополагающей своей задачей ставит разработку общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий. Корпус представляет собой большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач (определение В.П. Захарова) [10, с. 3].

Рассмотрим возможности лингвистических исследований русского языка средствами Национального корпуса русского языка ИРЯ им. В.В. Виноградова РАН (далее –

НКРЯ), доступ к которому осуществляется по адресу <http://www.ruscorgora.ru>. Научно-образовательный ресурс НКРЯ предоставляет пользователю свободный поиск предмета исследования в базе текстов (объемом более 600 млн слов) в нескольких подкорпусах. Отметим существенные характеристики и возможности каждого из них при проведении различных лингвистических исследований.

1. *Основной корпус* (основная часть НКРЯ) содержит нехудожественные и художественные, письменные и зафиксированные устные литературные тексты (середины XX – начала XXI в.), а также ранние тексты (XVIII – середины XX в.), снабженные морфологической разметкой, которая осуществляется с помощью специализированных программ автоматического анализа [18; 23]. Важной характеристикой данного подкорпуса является присутствие в некоторой части общего массива текстов снятой грамматической омонимии, что расширяет и упрощает возможности использования корпуса при различных морфологических и семантических исследованиях.

2. *Глубоко аннотированный (синтаксический) корпус*, в котором совместно с прописанной для слова морфологической информацией задана и синтаксическая структура предложения, в котором обнаруживается предмет поиска. Здесь в отличие от основного корпуса (морфологического) содержатся тексты со снятой морфологической и синтаксической омонимией [2]. Представленные в результате поиска структурные схемы предложений, синтаксические отношения и лексические функции способствуют применению средств поиска данного подкорпуса в синтаксическом и семантико-синтаксическом исследованиях.

3. *Газетный корпус* охватывает тексты как печатных, так и электронных СМИ с начала XX в. Подкорпусом газетного является корпус региональной и зарубежной прессы. Массив постепенно расширяется и может использоваться для различных стилистических исследований на материале публицистических текстов, а также лексических исследований, т. к. позволяет изучать языковые изменения «в режиме реального времени» средствами лексико-грамматического поиска.

4. *Корпус параллельных текстов*, или выровненный корпус, в котором собраны различные русские тексты и их перевод на другие языки (и наоборот), прошедшие процедуру соответствия. Процесс пользовательского поиска осуществляется по словоформе или слову,

а также средствами лексико-грамматического поиска. Данный массив широко используется в исследованиях по теории и практике перевода. Используются переводы на английский, белорусский, болгарский, бурятский, испанский, итальянский, китайский, латышский, немецкий, польский, украинский, французский, шведский, эстонский языки [9].

5. *Корпус диалектных текстов*, используемый диалектологами и специалистами по лингвистической географии, представляет собой массив записей диалектной речи различных регионов России. Здесь присутствуют толкования диалектных лексем, сохранены лексическая, морфологическая и синтаксическая особенности диалектов [13].

6. *Корпус поэтических текстов* составляют стихотворные тексты (на сегодняшний день не включены драматические). Наряду с морфологической и семантической специальная стиховедческая разметка позволяет проводить на базе данного массива литературоведческие исследования [7].

7. *Обучающий корпус русского языка*, по сравнению с остальными корпусами, представленными в информационно-справочной системе НКРЯ [22], количественно довольно небольшой и ориентирован в первую очередь на использование его средств при преподавании русского языка в школе.

8. *Корпус устной речи* составляет массив расшифрованных записей (в основном магнитофонных) частной и публичной устной речи с использованием традиционной орфографии. Возможен семантический, морфологический и лексический поиск, а также установки социологических параметров поиска [8]. Данный корпус предоставляет широкие возможности для изучения живой речи в динамике и исследований в области лексикологии, морфологии, семантики, лингвокультурологии, теории дискурса и т. д. Хронологический охват корпуса – 1900–2000-е гг.

9. *Акцентологический корпус (корпус истории русского ударения)* содержит следующие массивы текстов: базу поэтического корпуса, стихотворные тексты поэтов-любителей (разработчики корпуса отмечают, что «эстетические достоинства этих текстов не важны для исследования русской акцентологии, но регулярность чередований ударных и безударных слогов в русском стихе дает бесценный материал для уяснения множества вопросов, связанных с расстановкой ударений в словах современными носителями русского язы-

ка» [17]), а также записи звучащей устной речи и кинофильмов. Результатами поиска являются текстовые отрывки с проставленным в словах ударением, что предоставляет возможности для исследований в области акцентологии, стиховедения, а также в обучении русскому языку, в том числе и как иностранному.

10. *Мультимедийный корпус* состоит из фрагментов кинофильмов 1930–2000-х гг., представленных параллельной текстовой расшифровкой звучащей речи, аудио- и видеорядом, которые находятся в свободном доступе, в том числе при поисковой выдаче. Отмечаются также жесты, которые присутствуют на видео. Особенность подкорпуса в том, что поиск можно осуществлять не только по слову и средствами лексико-грамматического поиска (как и во всем НКРЯ), но и по жестам и речевым действиям, отраженным в корпусе.

11. *Древнерусский корпус*. Поиск с возможностью выбора орфографии (точной, упрощенной или модернизированной) осуществляется в массиве различных текстов на древнерусском языке, в который входят как оригинальные тексты, так и переводы с греческого и южнославянские письменные памятники, переписанные на Руси. Средства поиска и представление выдачи предназначены для широкого круга пользователей, а также для исторических исследований и работ по истории русского языка.

Таким образом, массив текстов НКРЯ представляет собой структурированную информационную систему, репрезентативность которой представлена в возможности выбора стилистической и жанровой принадлежности текстов исследования, их устной или письменной сферы бытования, времени создания и языка представления. Подкритериями выбора части корпусного массива являются грамматическая омонимия, основные параметры текста, временные версии (т. к. некоторые тексты обновляются), жанр и тип текста и орфография. Это делает возможным пользовательское создание собственного подкорпуса. На наш взгляд, данная функция является наиболее важной при использовании системы в научных лингвистических исследованиях. Рассмотрим методику поиска и структурирования языковых явлений средствами НКРЯ на примере исследования лексической антонимии, т. е. явления семантических отношений двух лексических единиц, имеющих противоположные значения.

Целью исследования определим характеристику функций лексических антонимов в

публицистических текстах современного русского языка. Одной из задач, таким образом, является поиск явлений в контекстах нехудожественных текстов современного русского языка. Выстроим алгоритм поиска.

1. Определение пользовательского подкорпуса
  - 1.1. Жанр и тип текста
  - 1.2. Орфография
2. Расстановка параметров лексико-грамматического поиска в пользовательском подкорпусе
  - 2.1. Первая лексема пары
    - 2.1.1. Грамматические признаки
    - 2.1.2. Семантические признаки
    - 2.1.3. Дополнительные признаки
    - 2.1.4. Словообразовательные признаки
  - 2.2. Вторая лексема пары
    - 2.2.1. Грамматические признаки
    - 2.2.2. Семантические признаки
    - 2.2.3. Дополнительные признаки
    - 2.2.4. Словообразовательные признаки
  - 2.3. Расстановка дистанции между членами пары
3. Получение необходимой выборки и сохранение необходимой для исследования информации

*Шаг 1. Определение пользовательского подкорпуса.* При выборе параметров для создания собственного подкорпуса следует, на наш взгляд, опустить выбор снятой / неснятой грамматической омонимии и основных параметров текста. Первое связано с тем, что большая часть текстов НКРЯ не несет разметку снятой грамматической омонимии, и, следовательно, результат при выбранном параметре *Только тексты со снятой грамматической омонимией* будет отличаться количественно от результатов поиска при упоминании или неупоминании обоих вариантов. Определение основных параметров текста сужает область до определенного произведения или автора, что тоже сказывается на репрезентативности результата. Поэтому в нашем алгоритме переходим сразу к шагу *1.1. Жанр и тип текста*. Ориентируясь на цель исследования, отмечаем выбор на **Нехудожественный текст**, далее выбираем сферу функционирования из открывающегося окна параметров (либо вводим вручную) – **публицистика**, и далее – тип текста, в котором необходим поиск антонимических пар (выбор или написание вручную через команду «|»). Затем переходим к шагу *1.2. Орфография*. В связи с тем, что исследование проводится на массиве текстов современного русского языка, ставим выбор на **Только тексты в новой орфографии**.

Далее система предлагает сохранить под-корпус и перейти к поиску. В данном разделе можно просмотреть все тексты, которые вошли в созданную пользователем базу.

Единицей поиска является пара лексем, которые мы будем брать из словаря антонимов Л.А. Введенской [5], т. к. в нем заглавная пара словарной статьи стилистически нейтральна и семантически однородна. При этом если слова имеют лексико-семантические варианты и антонимическую отличающуюся по словоформе пару к ним, то их следует считать новой единицей поиска. В связи с этим переходим сразу к *Шаг 2. Расстановка параметров лексико-грамматического поиска в пользовательском подкорпусе.*

Для поиска пары слов необходимы два поля. Выберем для примера антонимическую пару *жарко – холодно*. Вносим в поле *лексико-грамматического поиска «Слово»* лексему *жарко*. При необходимости можно выбрать различные грамматические признаки (2.1.1. *Грамматические признаки*). В нашем случае для более массовой выборки продублируем лишь категориальную принадлежность слова командой ADV или из диалогового окна – «наречие».

При выполнении шага 2.1.2. *Семантические признаки* важно отметить, что обе лексемы пары многозначны и, согласно словарю антонимов Л.А. Введенской, у лексико-семантических вариантов присутствуют свои антонимические пары (ЛСВ отмечаются индексом) [5, с. 142]:

| ЖАРКО <sub>1</sub> – ХОЛОДНО <sub>1</sub>                                         |                                                                                    |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| <i>В знач. сказ.</i> О высокой температуре окружающей среды (воздуха)             | <i>В знач. сказ.</i> О низкой температуре окружающей среды (воздуха)               |
| ЖАРКО <sub>2</sub> – ХОЛОДНО <sub>2</sub>                                         |                                                                                    |
| <i>В знач. сказ.</i> Об ощущении жары, высокой температуры, испытываемом кем-либо | <i>В знач. сказ.</i> Об ощущении холода, низкой температуры, испытываемом кем-либо |

По умолчанию поиск производится по семантической разметке всех словарных значений слова. Чтобы сузить область поиска, необходимо отметить один или два параметра: *1-е знач.* – поиск по семантической разметке первого словарного значения слова; *др. знач.* – поиск по семантической разметке прочих (переносных) значений слова: *фильтр 1* – поиск в контекстах с разрешенной лексико-семантической омонимией, где слово выступает в первом значении (например, если за-

дано прилагательное *мягкий* с семантическим признаком ‘физические свойства’, то будут найдены контексты типа *мягкий снег* или *мягкая ткань*, но не *мягкий человек*); *фильтр 2* – поиск в контекстах с разрешенной лексико-семантической омонимией, где слово выступает в других (переносных) значениях (таким образом, по запросу *мягкий + качества человека* будут найдены контексты типа *мягкий человек* или *мягкий большевик*). Комбинация параметров *1-е знач.* и *др. знач.* задает поиск по всем значениям слова. Комбинация параметров *фильтр 1* и *фильтр 2* задает поиск по всем контекстам со снятой лексико-семантической омонимией [17].

В случае поиска антонимов в первом значении можно отметить *1-е знач.*, во втором и возможных других – *др. знач.* Однако при этом могут пострадать количественная и качественная характеристики охвата результата, поскольку толкование слов в корпусе задано средствами различных толковых словарей, в которых возможен различный порядок фиксации лексико-семантических вариантов. В связи с этим следует выбрать параметры *1-е знач.* и *др. знач.* и в процессе анализа результатов особо учитывать наличие полисемии в выборке.

2.1.3. *Дополнительные признаки* устанавливаются в случае необходимости поиска слова в определенной позиции: перед знаком препинания и после него, в начале и конце предложения. В нашем случае не следует выбирать дополнительные условия. При выборе в шаге 2.1.4. *Словообразовательные признаки* действия те же – при отсутствии выбора по умолчанию в выборке отражаются все варианты.

Далее предпринимаем те же шаги со второй лексемой. Переходим к пункту 2.3. *Расстановка дистанции между членами пары*. Он служит для более точного определения структуры сочетания слов. Между соседними словами расстояние считается равным одному слову; расстояние, равное нулю, означает совпадение словоформ. Отрицательное число в поле «*Расстояние*» соответствует обратному порядку слов:

– условие «от –3 до 3» будет выполнено для пары слов, которые отстоят друг от друга не более чем на 3 словопозиции в любую сторону;

– условие «до –3» выберет потомков, отстоящих на расстояние от 0 до 3 слов *влево* от слова-родителя;

– условие «от –3» выберет потомков, стоящих на расстоянии 3 и более слов *влево* от слова-родителя [17].

Поиск в *расстояние 1*, например, возможен, когда требуется найти фигуры с точной языковой структурой, нарушение которой недопустимо. В нашем случае неважно расстояние, важно присутствие пары слова в контексте, поэтому не выставляем никаких значений во второе поле, в первое ставим «1».

3. *Получение необходимой выборки и сохранение необходимой для исследования информации.* Результат поиска открывается в новом окне с пометой «*жарко ADV на расстоянии 1 от холодно ADV. Найдено 24 документа, 24 вхождения*». На новой странице присутствуют 10 контекстов на страницу. Каждый результативный контекст (корпусный эквивалент традиционной карточки) представляет собой:

– заголовок (название текста, год, наличие или отсутствие разметки по наличию омонимии и количество найденных контекстов в этом тексте);

– сам контекст с выделенным цветом предметом поиска;

– библиографическую ссылку;

– упоминание о разметке по критерию «омонимия»;

– гиперссылку, которая ведет на отдельное окно с расширенным контекстом (чаще всего абзац).

Каждое слово в предоставляемом контексте выборки само по себе является гиперссылкой, при переходе на которую во всплывающем окне открывается таблица с разметкой параметров каждой единицы (грамматических, семантических и дополнительных) и ссылкой на ее фиксацию в электронных вариантах различных словарей.

Результаты выдачи сохраняются на персональный компьютер в форматах Excel, OpenOffice Calc, XML. Таким образом, оптимизация времени, затраченного на поиск речевых проявлений заданных антонимических пар, а также широта охвата и систематизация лингвистических характеристик выборки единиц открывают различные возможности для изучения их стилистических функций (в разметке контекстом присутствует помета «Стиль»), проведения структурных, синтаксических, семантических и других видов анализа. При использовании актуализированных текстов корпуса становится возможным изучение современного, «в реальном времени» со-

стояния языка. Открытость информационно-поисковой системы НКРЯ для пользователей и его научно-образовательная направленность расширяют возможности решения различных исследовательских задач в современной лингвистической науке.

### Список литературы

1. Авдеев Ю.А. Инновационные технологии в современной неологии: от картотеки к лингвистическому корпусу // Вестн. Адыг. гос. ун-та. Сер. 2: Филология и искусствоведение. 2010. №. 2. С. 87–90.

2. Апресян Ю.Д., Богуславский И.М., Иомдин Б.Л. [и др.]. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 193–214.

3. Баранов А.Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1–2. С. 179–191.

4. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. М.: Эдиториал УРСС, 2001.

5. Введенская Л.А. Словарь антонимов русского языка: более 500 антоним. гнезд. М.: Изд-во Астрель, 2003.

6. Ганева И.Ф. Об использовании корпусов в лингвистических исследованиях // Вестн. Башкир. ун-та. 2007. Т. 12. № 4. С. 104–106.

7. Гришина Е.А., Корчагин К.М., Плунгян В.А., Сичинава Д.В. Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 71–113.

8. Гришина Е.А., Савчук С.О. Корпус устных текстов в НКРЯ: состав и структура // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 129–149.

9. Добровольский Д.О., Кретов А.А., Шаров С.А. Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 263–296.

10. Захаров В.П. Корпусная лингвистика: учеб.-метод. пособие. СПб., 2005.

11. Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Национальный корпус русского языка как инструмент семантико-грамматического исследования лексики // Корпусная лингвистика – 2004: Междунар. конф.: тез. докл. СПб.: СПбГУ, 2004. С. 50–51.

12. Кутузов А.Б. Корпус несовершенных переводов: необходимость проекта // Проблемы тео-

- рии, практики и дидактики перевода: сб. науч. тр. Н. Новгород: НГЛУ, 2011. Вып. 14. Т. 1. С. 51–54.
13. Летучий А.Б. Корпус диалектных текстов: задачи и проблемы // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 215–232.
14. Мамонтова В.В. Корпусная лингвистика и лингвистические корпусы // Язык. Текст. Дискурс: науч. альманах Ставропольского отделения РАЛК / под ред. проф. Г.Н. Манаенко. Ставрополь: Изд-во ПГЛУ, 2007. Вып. 5. С. 275–283.
15. Михайлов М. Н. Компьютерное обеспечение корпуса текстов (взгляд пользователя) // Русистика сегодня. 1998. № 1–2. С. 192–201.
16. Нагель О.В. Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // Язык и культура. 2008. № 4. С. 53–59.
17. Национальный корпус русского языка [Электронный ресурс]. URL: <http://ruscorpora.ru/index.html> (дата обращения: 10.03.2018).
18. Оскольская С.А. Корпус письменных текстов XIX века: сферы употребления и жанровое разнообразие // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 46–51.
19. Падучева Е.В. НКРЯ как ресурс при исследовании предметной соотнесенности имен // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 374–382.
20. Перцов Н.В. О роли корпусов в лингвистических исследованиях // Корпусная лингвистика – 2006: тр. Междунар. конф. СПб.: Изд-во С.-Петерб. ун-та, 2006. С. 318–331.
21. Поляков А.Е. Технология подготовки информации в Национальном корпусе русского языка // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М., 2005. С. 175–192.
22. Савчук С.О., Сичинава Д.В. Обучающий корпус русского языка и его использование в преподавательской практике // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 317–334.
23. Савчук С.О. Корпус текстов первой половины XX века: текущее состояние и перспективы // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 27–45.
- \* \* \*
1. Avdeev Ju.A. Innovacionnye tehnologii v sovremennoj neologii: ot kartoteki k lingvisticheskomu korpusu // Vestn. Adyg. gos. un-ta. Ser. 2: Filologija i iskusstvovedenie. 2010. № 2. S. 87–90.
2. Apresjan Ju.D., Boguslavskij I.M., Iomdin B.L. [i dr.]. Sintaksicheski i semanticheski annotirovannyj korpus russkogo jazyka: sovremennoe sostojanie i perspektivy // Nacional'nyj korpus russkogo jazyka: 2003–2005. М.: Indrik, 2005. S. 193–214.
3. Baranov A.N. Avtomatizacija lingvisticheskikh issledovanij: korpus tekstov kak lingvisticheskaja problema // Rusistika segodnja. 1998. № 1–2. S. 179–191.
4. Baranov A.N. Vvedenie v prikladnuju lingvistiku: ucheb. posobie. М.: Jeditorial URSS, 2001.
5. Vvedenskaja L.A. Slovar' antonimov russkogo jazyk: bolee 500 antonim. gnezd. М.: Izd-vo Astrel', 2003.
6. Ganeva I.F. Ob ispol'zovanii korpusov v lingvisticheskikh issledovanij // Vestn. Bashkir. un-ta. 2007. T. 12. № 4. S. 104–106.
7. Grishina E.A., Korchagin K.M., Plungjan V.A., Sichinava D.V. Pojeticheskij korpus v ramkah NKRJa: obshhaja struktura i perspektivy ispol'zovanija // Nacional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 71–113.
8. Grishina E.A., Savchuk S.O. Korpus ustnyh tekstov v NKRJa: sostav i struktura // Nacional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 129–149.
9. Dobrovol'skij D.O., Kretov A.A., Sharov S.A. Korpus parallel'nyh tekstov: arhitektura i vozmozhnosti ispol'zovanija // Nacional'nyj korpus russkogo jazyka: 2003–2005. М.: Indrik, 2005. S. 263–296.
10. Zaharov V.P. Korpusnaja lingvistika: ucheb.-metod. posobie. SPb., 2005.
11. Kustova G.I., Ljashevskaja O.N., Paducheva E.V., Rahilina E.V. Nacional'nyj korpus russkogo jazyka kak instrument semantiko-grammaticeskogo issledovanija leksiki // Korpusnaja lingvistika – 2004: Mezhdunar. konf.: tez. dokl. SPb.: SPbGU, 2004. S. 50–51.
12. Kutuzov A.B. Korpus nesovershennyh perevodov: neobhodimost' proekta // Problemy teorii, praktiki i didaktiki perevoda: sb. nauch. tr. N. Novgorod: NGLU, 2011. Vyp. 14. T. 1. S. 51–54.
13. Letuchij A.B. Korpus dialektnyh tekstov: zadachi i problemy // Nacional'nyj korpus russkogo jazyka: 2003–2005. М.: Indrik, 2005. S. 215–232.
14. Mamontova V.V. Korpusnaja lingvistika i lingvisticheskie korpusy // Jazyk. Tekst. Diskurs: nauch. al'manah Stavropol'skogo otdelenija RALK / pod red. prof. G.N. Manaenko. Stavropol': Izd-vo PGLU, 2007. Vyp. 5. S. 275–283.
15. Mihajlov M.N. Komp'juternoe obespechenie korpusa tekstov (vzgljad pol'zovatelja) // Rusistika segodnja. 1998. № 1–2. S. 192–201.
16. Nagel' O.V. Korpusnaja lingvistika i ee ispol'zovanie v komp'juterizirovannom jazykovom obuchenii // Jazyk i kul'tura. 2008. № 4. S. 53–59.

17. Nacional'nyj korpus ruskogo jazyka [Elektronnyj resurs]. URL: <http://ruscorpora.ru/index.html> (data obrashhenija: 10.03.2018).

18. Oskol'skaja S.A. Korpus pis'mennyh tekstov XIX veka: sfery upotreblenija i zhanrovoe raznoobrazie // Nacional'nyj korpus ruskogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 46–51.


19. Paducheva E.V. NKRJa kak resurs pri issledovanii predmetnoj sootnesennosti imen // Nacional'nyj korpus ruskogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 374–382.

20. Percov N.V. O roli korpusov v lingvisticheskih issledovanijah // Korpusnaja lingvistika – 2006: tr. Mezhdunar. konf. SPb.: Izd-vo S.-Peterb. un-ta, 2006. S. 318–331.

21. Poljakov A.E. Tehnologija podgotovki informacii v Nacional'nom korpusе ruskogo jazyka // Nacional'nyj korpus ruskogo jazyka: 2003–2005. Rezul'taty i perspektivy. M., 2005. S. 175–192.

22. Savchuk S.O., Sichinava D.V. Obuchajushhij korpus ruskogo jazyka i ego ispol'zovanie v prepodavatel'skoj praktike // Nacional'nyj korpus ruskogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 317–334.

23. Savchuk S.O. Korpus tekstov pervoj poloviny XX veka: tekushhee sostojanie i perspektivy // Nacional'nyj korpus ruskogo jazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istorija, 2009. S. 27–45.



***National corpus of the Russian language as the main search tool in linguistic research (based on the search for antonyms in journalistic texts)***

*The article deals with the implementation of various linguistic research by means of the national corpus of the Russian language. As the example of search for necessary units optimized in time and scope of the text material, it presents the algorithm of search for the pairs of antonyms in journalistic texts.*


**Key words:** *linguistic research, corpus, research material, national corpus of the Russian language, antonyms, search algorithm.*

(Статья поступила в редакцию 15.03.2018)

**Р.Ф. АБДУЛЛАЕВ, И.А. ДЕМИДОВА**  
(Калининград)

**ЯДЕРНЫЕ КОМПОНЕНТЫ  
МИКРОПОЛЯ ОБЪЕКТИВНО-  
СУБЪЕКТИВНОЙ НЕОБХОДИМОСТИ  
(гендерный аспект)**

*Рассматривается специфика функционирования основных экспликаторов модального значения необходимости (наречий «надо», «нужно», предикативного наречия «должен / должна») в извлеченных из газетно-публицистических текстов высказываниях мужчин и женщин. Выявляется сравнительная частотность употребления указанных экспликативов в речи мужчин и женщин, устанавливаются причины соответствующих различий, связанные с целеустановкой речевой коммуникации и гендерными факторами.*



**Ключевые слова:** *модальность, модальность необходимости, гендер, гендерная лингвистика, речевая коммуникация, газетный текст.*

В современной лингвистике модальность относится к важнейшим семантическим категориям, которые устанавливают непосредственную связь высказывания с внеязыковой действительностью и формируют его коммуникативный потенциал. Непрерывающийся интерес к языковой модальности является закономерным (в особенности к межкатегориальным отношениям модальности (см., например: [2–4; 14])) на основе междисциплинарного подхода с привлечением фактических данных из различных гуманитарных наук: социолингвистики, психолингвистики, гендерной лингвистики и др. Междисциплинарный подход позволяет более глубоко понять роль модальности в речевой коммуникации [11, с. 17].

В контексте междисциплинарного подхода весьма актуальным представляется исследование ранее детально не рассматривавшейся категории модальности в русле гендерных исследований – особенности выражения модальных значений в гендерном аспекте. Стоит отметить, что обращение к категории модальности «позволяет понять гендерные закономерности в модальности как коммуникативной категории, выявить механизм использования гендерно обусловленных коммуникативных стратегий, а также установить природу гендерных стереотипов» [1, с. 9].