Вестник Томского государственного университета. 2019. № 446. С. 29–40. DOI: 10.17223/15617793/446/4

УДК 81’33

К.С. Кочергина

ЛИНГВИСТИЧЕСКАЯ БАЗА ДАННЫХ ОТРИЦАТЕЛЬНО-ОЦЕНОЧНОЙ ЛЕКСИКИ: КОНЦЕПЦИЯ, СТРУКТУРА, НАПОЛНЕНИЕ

Исследование проведено при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере по программе «УМНИК» по теме «Разработка программного обеспечения для поддержки процедуры лингвистической экспертизы» в рамках договора № 8038ГУ/2015 от 23.11.2015 г.

Приводится обзор существующих на данный момент лингвистических баз данных и выявляются основания для их классификации. Предлагается концепция разрабатываемой лингвистической базы данных отрицательно-оценочной лексики (имен существительных со значением номинации лица и группы лиц), описывается ее структура, демонстрируется ее наполнение на примере первого тома «Русского семантического словаря» под ред. Н.Ю. Шведовой (1998). Рассматриваются возможности ее применения в лингвистических исследованиях и других видах деятельности.

Ключевые слова: лингвистическая база данных; имя существительное; отрицательно-оценочная лексика; стилистическая помета; словарь; юридическая лингвистика; лингвистическая экспертиза.

Постановка проблемы

При решении теоретических и прикладных задач в лингвистике может возникнуть необходимость обращения к нескольким лексикографическим источникам, например, при производстве лингвистической экспертизы, когда разнородность систем стилистических помет и расхождение в стилистической квалификации лексики в словарях современного русского языка приводят к снижению точности и объективности проводимых с опорой на них исследований [1]. Одним из решений указанной проблемы может стать применение технологии базы данных, позволяющей получать лексикографическую информацию из различных источников в сопоставлении.

Актуальность данного исследования также обусловлена потребностью юридической лингвистики в специализированном инструменте – лексикографическом издании, пригодном для целей лингвистической экспертизы [2], и необходимостью совершенствования информационных технологий в практике производства лингвистической экспертизы [3] в связи с возрастающей ролью информационных технологий в гуманитарных исследованиях и отсутствием опыта применения таковых в данной области. Представление такого лексикографического продукта в виде базы данных позволит реализовать новый подход к проведению лингвистических экспертиз – с использованием средств автоматизации [4].

База данных (БД) – это «упорядоченная совокупность данных, предназначенных для хранения, накопления и обработки с помощью ЭВМ» [5. С. 96]. Под лингвистической БД (ЛБД), соответственно, понимается совокупность систематизированных лингвистических данных.

Цель настоящей статьи заключается в описании концепции, структуры и наполнения ЛБД отрицательно-оценочной лексики, разрабатываемой для проведения лингвистических исследований, в частности, лингвистической экспертизы.

Источниками данной ЛБД являются словари современного русского языка (конца XX – начала XXI в.), при отборе которых учитываются методические рекомендации по производству лингвистической экспертизы и практика применения словарей в качестве лексикографических источников для нее (по данным экспертных заключений). Наполнение разрабатываемой ЛБД представлено на примере первого тома «Русского семантического словаря» под ред. Н.Ю. Шведовой (1998) [6. Т. 1].

Материалом для исследования послужили имена существительные со значением номинации лица и группы лиц и отрицательно-оценочной семантикой, отобранные из указанного словаря методом сплошной выборки. Отбор материала проведен с опорой на идеографический, семантический, стилистический принципы. В исследовании также применен описательно-аналитический (общенаучный) метод с совокупностью его приемов. При разработке ЛБД использован метод концептуального проектирования, состоящий в определении параметров описания единиц информации и связей между ними в соответствии с задачами, для решения которых создается информационно-справочный ресурс.

История вопроса

Технология БД широко применяется в различных сферах, в том числе лингвистике, например, «в типологических и сравнительных, фонетических, лексических, грамматических исследованиях; лексикографии; семантических исследованиях, включая составление тезаурусов; в компьютерной лингвистике и для решения прикладных лингвистических задач (лингводидактических, автоматизированного перевода, автоматического распознавания и синтеза речи)» [7. С. 26]. В работах Н.А. Мишанкиной рассматриваются проблемы, связанные с обработкой информации и созданием информационных систем на основе технологии БД в гуманитарной сфере и, в частности, с применением данной технологии в области филологических (лингвистических) исследований [7, 8]. Исследователь дает историческую справку о существующих БД и предлагает выделять ЛБД в отдельную категорию, подразделяя ее на два типа: *полнотекстовые* (содержащие полные тексты) и *фактографи-*

ческие (содержащие лингвистические единицы различных языковых уровней).

В качестве примера полнотекстовой ЛБД назовем разработанную коллективом Томского государственного университета (Н.А. Мишанкиной, И.В. Тубаловой, Ю.А. Эмер) ***БД «Среднеобский фольклор»*** [8, 9]. Целью создания этой БД является объединение и представление ранее недоступного фольклорного материала народной культуры Сибири в доступном и удобном формате для различных исследований в области антропологии, этнографии, культурологии, лингвистики, истории сибирского региона. В эту БД входят фольклорные тексты на русском языке в различных жанрах, например частушки, загадки, пословицы, считалки, скороговорки.

К первому типу ЛБД также можно отнести *электронные библиотеки*, *текстовые коллекции* и *корпуса текстов*, например, «Национальный корпус русского языка» (ruscorpora.ru), «Томский диалектный корпус» (http://losl.tsu.ru/?q=corpus/demo).

Среди ЛБД второго типа по уровню описания в них языка можно выделить *фонетические*, *морфологические*, *лексические* и *синтаксические* ЛБД соответственно.

ЛБД, в которых представлены лексические единицы, особенно активно применяются в лексикографической практике для составления как традиционных, так и электронных словарей. А.Н. Баранов сравнивает две технологии создания словаря (традиционную и компьютерную), отмечая явное преимущество последней: «Традиционная форма фиксации словарных данных – каталожная карточка, в которой указывается описываемое слово, пример употребления, источник примера, автор, а также различная дополнительная информация. Современные компьютерные технологии позволяют упростить процесс сбора и хранения лексикографической информации, используя вместо обычной картотеки базу данных, записи которой представляют собой аналог традиционной каталожной карточки» [10. С. 82]. Однако лексические ЛБД используются не только при создании новых словарей, но и применяются в различных лингвистических исследованиях, поскольку ЛБД как самостоятельный формат представления лексикографической информации обладает рядом преимуществ перед словарями: позволяет накапливать, хранить, систематизировать лингвистическую информацию и осуществлять быстрый поиск в ее пределах, а также имеет открытую структуру для внесения изменений, непрерывно происходящих в языковой системе [11].

Рассмотрим известные на данный момент проекты фактографических ЛБД, содержащих единицы лексического уровня и ориентированных на описание лексической семантики.

БД «Звучание» (Н.А. Мишанкина, Томский государственный университет) включает глаголы, содержащие в своей семантике компонент «звук», отражает лексико-семантическое поле «Звучание» и предусматривает возможность пополнения аудио- или видеоматериалами, демонстрирующими тип звучания, его субъект и связанное с ним действие [7].

БД метафорической терминологии ориентирована на выявление универсальных метафорических моделей, отражающих концептуальные метафоры, являющиеся ключевыми для русского научного дискурса. Авторы (Н.А. Мишанкина, Е.А. Панасенко, Томский политехнический университет) указывают, что разрабатываемая БД может применяться для решения ряда частных задач, например, для выявления наиболее востребованных в терминообразовании лексических единиц национального языка. Материалом для создания этой БД послужили терминологические словари различных научных областей: биологии, геологии, информатики, лингвистики, медицины, психологии, социологии, философии, экологии. Общее количество проанализированных источников составило 65, а количество выявленных метафорических терминологических единиц – около 2 000. Центральной таблицей этой БД является «Лексическая единица», вспомогательными – «Метафорическая модель» и таблицы предметных областей, отражающие метафорическую терминологию конкретной предметной области [12].

Психолингвистическая БД (проект Томского государственного университета: З.И. Резановой, А.А. Миклашевского) используется для описания образно-перцептивного компонента языковой семантики и позволяет применять статистические методы для анализа слабоструктурированной семантической информации. Эта БД содержит полученную в ходе проведения психолингвистического эксперимента информацию о 506 как конкретных, так и абстрактных существительных, а именно об их связи с модальностями восприятия (зрением, слухом, вкусовыми, обонятельными и осязательными ощущениями) и другую дополнительную информацию, например, о субъективном возрасте усвоения слова [13].

БД прагматически маркированной лексики разрабатывается коллективом Новосибирского государственного педагогического университета: Т.А. Трипольской, Е.Ю. Булыгиной, В.Д. Черняк (г. Санкт-Петербург), В.А. Ефремовым (г. Санкт-Петербург), М.А. Лаппо, Е.Г. Басалаевой, Л.Н. Храмцовой, И.И. Саженины. Демонстрационная версия этой БД, включающая в себя около 400 единиц, доступна в интернете (spml.ipmip.nspu.ru). Согласно представленному в статьях [14, 15], описанию БД формально и содержательно делится на две части. В первой части представлены словарные статьи следующих классических и современных словарей русского языка: «Толковый словарь русского языка» под ред. Д.Н. Ушакова (1935–1940), «Словарь современного русского литературного языка» в 17 т. (1950–1965), «Словарь русского языка» в 4 т. под ред. А.П. Евгеньевой (1981–1984), «Толковый словарь русского языка» С.И. Ожегова и Н.Ю. Шведовой (1999), «Толковый словарь русского языка конца XX века. Языковые изменения» под ред. Г.Н. Скляревской (1998), «Большой толковый словарь русского языка» под ред. С.А. Кузнецова (1998). Вторая часть БД представляет собой словарные статьи «Словаря прагматически маркированной лексики», разработанного авторами-составителями проекта, и включает в себя толкование,

иллюстративный материал и подробное описание прагматической зоны значения слова (эмоциональная оценка, идеологический, гендерный, социально-статусный, возрастной, национально-культурный компоненты) с опорой на материалы «Национального корпуса русского языка». Более подробно структура *БД прагматически маркированной* лексики описывается в работе [16], где также представлены принципы, лежащие в основе создаваемой БД, критерии отбора языковых единиц, их лексикографическая интерпретация, разработанная система помет, типизация лексикографического комментария. Здесь описывается поисковая система данной БД, при этом техническая сторона ее создания характеризуется в работе [17], в которой дается описание двух таблиц БД: первой, хранящей заголовки словарных статей, и второй, содержащей собственно тексты словарного описания.

Словарная БД, фрагмент которой доступен в сети Интернет (ruslex-encode.ru), создается в Институте лингвистических исследований РАН (Е.В. Пурицкая, Д.И. Панков). В основе этой БД лежат данные из «Большого академического словаря русского языка» (2004–2014, издание продолжается), а также других академических словарей, из которых извлекаются функционально-стилистические, специальные, динамические пометы при словах с момента их первой фиксации. *Словарная БД* предназначена для предоставления справочной информации о стилистической характеристике лексики в ее историческом развитии [18].

Проведенный анализ позволяет сделать вывод о том, что каждая из рассмотренных ЛБД создается для решения конкретной исследовательской задачи, и следовательно, они различаются по ряду критериев, на основании чего можно выделить следующие подтипы ЛБД:

– *по типу лексических единиц*, а именно по их *частеречной принадлежности*: ЛБД, содержащие глаголы (*БД «Звучание»*), имена существительные (*психолингвистическая БД*, *БД прагматически маркированной лексики*), единицы нескольких частей речи (*словарная БД*); а также по их *сфере употребления*: ЛБД, содержащие специальную лексику (*БД метафорической терминологии*);

– *по аспекту описания лексической семантики*: ЛБД, в которых описывается семантический компонент «звук» (*БД «Звучание»*), метафорическая семантика (*БД метафорической терминологии*), образно-перцептивный компонент (*психолингвистическая БД*), прагматический макрокомпонент (*БД прагматически маркированной лексики*), нормативно-стилистический компонент в динамическом аспекте (*словарная БД*);

– *по источнику данных*: ЛБД, содержащие данные нескольких типов (*БД «Звучание»*), экспериментальные данные (*психолингвистическая БД*), словарные данные (*БД метафорической терминологии*, *БД прагматически маркированной лексики*, *словарная БД*);

– *по области применения*: ЛБД, предназначенные для различных исследовательских целей (*БД «Звучание»*, *БД прагматически маркированной лексики*), исследования терминологии (*БД метафорической терминологии*), психолингвистических исследований (*психолингвистическая БД*), применения в лексикографической практике (*словарная БД*).

В рамках данного исследования также проведен поиск ЛБД, зарегистрированных в Федеральном институте промышленной собственности (new.fips.ru/iiss), результаты которого свидетельствуют об отсутствии не только ЛБД отрицательно-оценочной лексики, но и специализированных ЛБД, предназначенных для целей лингвистической экспертизы.

***

Таким образом, одним из продуктивных направлений на пересечении прикладной лингвистики и современной лексикографии является разработка ЛБД, применение которых позволяет как решать ряд новых, так и повышать эффективность решения традиционных для лингвистики задач. Использование технологии БД в лингвистических (и в целом – гуманитарных) исследованиях способствует не только удобному представлению материалов и их объединению в единую структуру, но и повышению эффективности работы с ними за счет возможности использования средств машинной обработки, а также открывает новые перспективы для дальнейших исследований на основе созданных БД или с их использованием.

1. Концепция лингвистической базы данных отрицательно-оценочной лексики.

Представим разрабатываемую в рамках данного исследования фактографическую ЛБД, содержащую лексические единицы – ***«Лингвистическую базу данных отрицательно-оценочной лексики (имена существительные со значением номинации лица и группы лиц)»*** [19] (далее – ЛБД ООЛ), основные принципы проектирования которой описаны в [20].

ЛБД ООЛ представляет собой совокупность систематизированных лингвистических данных – лексико-семантических единиц (ЛСЕ) и их параметров. Под ЛСЕ понимается лексическая единица в конкретном значении, а под ее параметрами – значимые семантические, стилистические и другие характеристики.

ЛБД ООЛ разрабатывается для применения в практике лингвистической экспертизы, объектом которой являются конфликтные тексты трех видов: тексты-неудачи, тексты-манипуляторы, тексты-злоупотребления (в терминологии О.Н. Матвеевой [21]). Для текстов-злоупотреблений (собственно конфликтных текстов в узком смысле) характерно наличие оценочных лексических единиц, выражающих отрицательное отношение субъекта оценки к объекту оценки, которым, как правило, является лицо и группа лиц.

Исходя из этого, на данный момент в ЛБД ООЛ включены ЛСЕ с отрицательно-оценочной семантикой, представляющие собой имена существительные (в том числе эквивалентные им единицы) со значением номинации лица и группы лиц. В фокусе данного исследования находятся имена существительные, поскольку, несмотря на то, что семантика оценки может быть выражена любой частью речи, именно существительные «характеризуют объект, выявляя основание оценки и делая оценочное суждение более кате-

горичным: эмоциональные оценки непосредственно приписываются объекту, а не передаются через характеристику его действий или свойств. Глагол, например, относит оценочную характеристику к определенному временному промежутку, прилагательное переносит акцент с лица на его свойства» [22. С. 132]. В перспективе возможно ее пополнение ЛСЕ других частей речи (например, глаголами со значением действия и состояния лица, именами прилагательными со значением свойства и качества лица), а также ЛСЕ с другими семантическими компонентами, например, положительно-оценочным.

В ЛБД ООЛ содержится семантическая и стилистическая (отрицательно-оценочная) информация о ЛСЕ, зафиксированная в толковании и при помощи стилистических помет (СП), извлеченная из лексикографических источников. Помимо словарных данных в ЛБД ООЛ содержится определенная на их основе информация о принадлежности ЛСЕ к тематическим группам и о возможности охарактеризовать форму негативной оценки, выраженной при помощи ЛСЕ, как неприличную.

Одним из принципов составления ЛБД ООЛ является опора на авторитетные (в том числе отобранные с учетом методических рекомендаций и практики производства лингвистической экспертизы) лексикографические издания, поскольку именно этот вид источников признается наиболее надежным [23] и единственным легитимным [24] для применения в экспертной практике.

На данный момент ЛБД ООЛ включает одно наименование лексикографического источника – толково-идеографический словарь современного русского литературного языка – первый том «Русского семантического словаря» (РСС) под ред. Н.Ю. Шведовой (1998). При этом представляется необходимым добавление в ЛБД ООЛ данных из словарей разных типов (об этом см., например, в [25, 26]). Также предполагается, что при достаточном обосновании в ЛБД ООЛ могут быть включены данные из источников другого вида, например, результаты экспериментальных исследований обыденного понимания неприличной формы.

Отметим, что представление данных из нескольких источников в формате ЛБД позволяет при проведении лингвистических исследований (как теоретических, так и практических, в том числе при производстве лингвистической экспертизы) не только отображать информацию из каждого источника в более удобном (структурированном) для анализа виде, но и сопоставлять ее из нескольких, тем самым объективируя результаты исследований.

ЛБД ООЛ планируется разместить в открытом доступе в интернете, возможно использование ЛБД ООЛ в виде компьютерной программы (приложения), а также подключение ЛБД ООЛ в качестве модуля к программам автоматической обработки текстов. На сайте и в программе (приложении) будут доступны два варианта работы пользователя с ЛБД ООЛ через удобный, интуитивно понятный интерфейс: поиск ЛСЕ в ЛБД ООЛ по введенному пользователем в поисковую строку заголовочному слову и поиск (сортировка, фильтрация, группировка) по выбранным пользователем параметрам, например, по алфавиту (по аналогии со словарем), тематическим группам, стилистическим пометам. Так, выбрав в качестве поискового параметра СП *бран*., пользователь получит список представленных в ЛБД ООЛ ЛСЕ, маркированных соответствующей СП. Особенность данной ЛБД заключается также в том, что она может быть применена для автоматизированного поиска отрицательно-оценочной лексики в конфликтных текстах [27], что особенно актуально для исследования текстов большого объема.

Содержащиеся в ЛБД ООЛ данные представляют практический интерес в первую очередь для лингвистов-исследователей, изучающих языковые явления на материале конфликтных текстов, и лингвистов-экспертов. Например, ЛБД ООЛ может применяться как инструмент исследования при решении вопросов о негативной оценке лица и группы лиц по делам об оскорблении, оскорблении представителя власти, неуважении к суду, оскорблении чувств верующих (поскольку содержит информацию о возможности охарактеризовать форму негативной оценки, выраженной при помощи ЛСЕ, как неприличную), а также по делам о клевете, по делам, связанным с экстремистской деятельностью (поскольку содержит информацию о принадлежности ЛСЕ к тематическим группам) и др. Кроме того, ЛБД ООЛ может быть востребована как информационно-справочный ресурс не только специалистами и экспертами, но и следователями, прокурорами, судьями, адвокатами и др.

***

Таким образом, в соответствии с подтипами ЛБД (выделенными выше при описании известных на данный момент проектов), ЛБД ООЛ можно охарактеризовать следующим образом: *по типу лексических единиц*, а именно по их *частеречной принадлежности* – ЛБД ООЛ содержит имена существительные; *по аспекту описания лексической семантики* – в ЛБД ООЛ описывается отрицательно-оценочный компонент; *по источнику данных* – ЛБД ООЛ содержит данные нескольких типов (словарные данные и определенную на их основе информацию: о тематических группах, к которым относятся ЛСЕ, и о возможности охарактеризовать форму выраженной при помощи ЛСЕ негативной оценки как неприличную); *по области применения* – ЛБД ООЛ предназначена для применения в лингвоэкспертной практике, а также может быть использована для различных исследовательских и практических целей.

2. Структура лингвистической базы данных отрицательно-оценочной лексики.

«Для создания и ведения Б[аз] д[анных] (обновления, обеспечения доступа к ним по запросам и выдачи их пользователю)» [5. С. 96] используются специализированные программы – системы управления базами данных (СУБД). А.Н. Баранов отмечает, что в настоящее время нет СУБД, предназначенных для лингвистических целей, однако для них пригодны существующие, например, D-Base, ACCESS, FOX-

Base, PARADOX [10. С. 82]. Настоящая ЛБД выполнена в реляционной СУБД Microsoft Access 2016, входящей в состав пакета Microsoft Office. Выбор данной программы обусловлен ее удобным графическим интерфейсом и наличием широкого спектра возможностей.

БД состоят, как правило, из нескольких взаимосвязанных таблиц (например, *БД метафорической терминологии* включает в себя три таблицы: одну основную и две вспомогательные), в которых содержатся данные, распределенные по строкам (записям) и столбцам (полям). Каждая запись представляет собой совокупность значений нескольких полей, которые в зависимости от своей функции содержат данные разных типов, например, текстовые, числовые, гиперссылки, вложения.

ЛБД ООЛ состоит из 11 таблиц данных: основная таблица семантико-стилистической характеристики отрицательно-оценочных слов в словаре, например, *Semantic_Stylistic_RSS* для РСС, и 10 таблиц, относящихся к вспомогательным и обеспечивающих связность и целостность данных в основных таблицах: таблица лексикографических источников *Dictionaries _List*; таблица отрицательно-оценочных лексических единиц *All_Lexical_Units_List*; таблица типов лексических единиц *Lexical_Units_Types*; таблица отрицательно-оценочных СП *Stylistic_Labels_List*; таблица отрицательно-оценочных СП в словаре, например, *Stylistic_Labels_RSS* для РСС; таблица хронологических помет *Chrono_Labels_List*; таблица типов лексического значения *Mean_Types_List*; таблица лексических классов *Lexical_Classes_List*; таблица тематических групп *Thematic_Groups_List*; таблица характеристики формы негативной оценки как неприличной *Indecent_Form*.

При этом основные таблицы семантико-стилистической характеристики отрицательно-оценочных слов в словаре и вспомогательные таблицы отрицательно-оценочных СП в словаре создаются для каждого последующего лексикографического источника, данные из которого включаются в ЛБД ООЛ, а такие таблицы, как *Dictionaries_List, All_ Lexical_Units_List, Lexical_Units_Types, Lexical_Classes_ List, Thematic_Groups_List,* являются таблицами открытого типа, т.е. пополняются записями и полями при добавлении в ЛБД ООЛ новых данных, в том числе из других источников.

По аналогии со словарем в ЛБД можно выделить макроструктуру, которая представляется в виде схемы данных, отображающей связи между таблицами, и микроструктуру – структуру каждой таблицы данных. Отметим, что для ЛБД, содержащих лексические единицы и извлеченную из лексикографических источников информацию о них, а также предназначенных для применения в лексикографической практике, характерна такая структура таблиц, которая соотносима со структурой словарной статьи. Так, например, в ЛБД ООЛ поле, содержащее ЛСЕ, соотносится с так называемой «левой частью» словарной статьи (зоной заглавного слова), а ряд других полей формирует так называемую «правую часть» словарной статьи (зоны помет, толкования, комментариев).

На рис. 1 приведена схема данных (логическая модель), на которой представлены все таблицы ЛБД ООЛ и при помощи линий отображены связи между ними (связи идут от обозначенных как «1» полей к обозначенным как «∞» полям), а также знаком «+» отмечены те поля, которые могут содержать более одного значения.

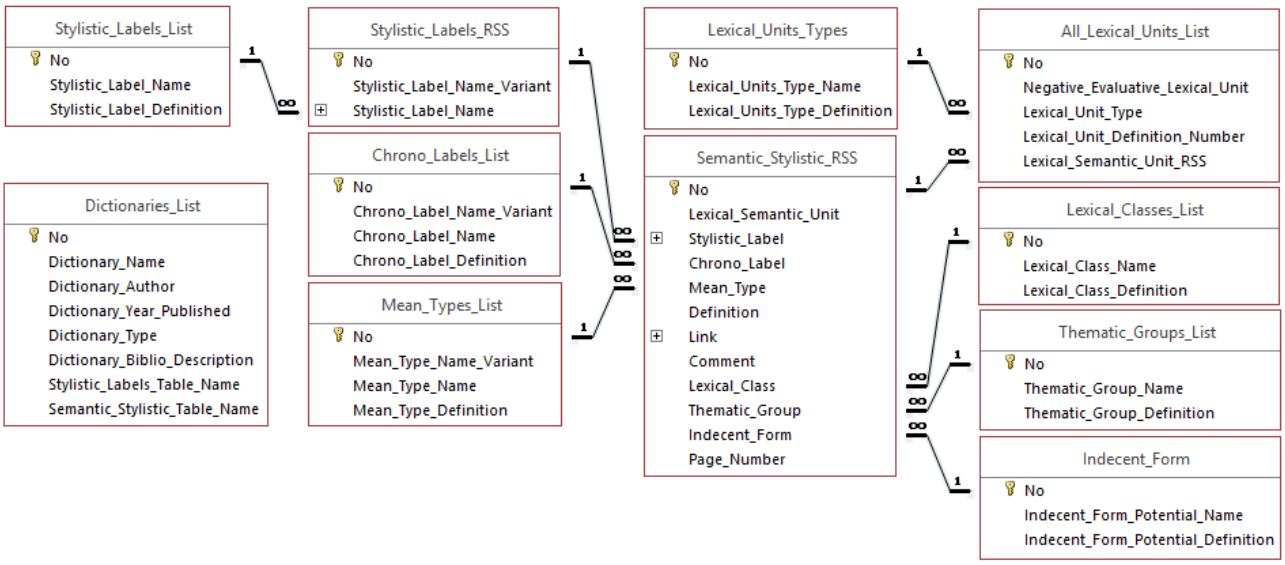

Рис. 1. Схема данных ЛБД ООЛ

В таблице ***Dictionaries_List*** зафиксирован список лексикографических источников, информация из которых включена в ЛБД ООЛ.

Таблица ***All_Lexical_Units_List*** представляет собой сводный перечень всех лексических единиц с отрицательно-оценочной семантикой (на данный момент – со значением номинации лица и группы лиц) и их значений, зафиксированных во всех словарях-источниках и представленных отдельными записями в ЛБД ООЛ в качестве ЛСЕ.

В таблице ***Lexical_Units_Types*** перечислены типы лексических единиц, например, имя существительное, устойчивое сочетание (фразеологическое выражение, идиома), поговорка, пословица.

Таблица ***Stylistic_Labels_List*** создана в виде списка отрицательно-оценочных СП с приведением толкования их значения (подробнее об отборе СП, маркирующих отрицательно-оценочную семантику, см. в [1]). Кроме того в данную таблицу включены кванторы типа *также, иногда, часто, чаще, теперь*, сочетание СП с которыми уточняет стилистическую (в частности, отрицательно-оценочную) характеристику слов.

В таблицах отрицательно-оценочных СП, формируемых для каждого словаря-источника ЛБД ООЛ, (например, в таблице ***Stylistic_Labels_RSS*** для РСС) приводятся отрицательно-оценочные СП в том виде, в котором они представлены в словарях-источниках, поскольку в них отмечается наличие разных вариантов условных обозначений одной и той же отрицательно-оценочной СП, а также встречаются комбинированные СП (например, образованные при помощи дефиса), СП в скобках, СП в сочетании с кванторами (подробнее об особенностях стилистической квалификации слов в словарях см. в [1]).

Таблица ***Chrono_Labels_List*** содержит варианты хронологических помет, используемых для временно́й характеристики слов в словарях (подробнее об этом см., например, в [1]).

В таблице ***Mean_Types_List*** содержится информация о типах лексического значения: прямом и переносном.

Таблица ***Lexical_Classes_List*** отражает лексические классы, к которым относятся ЛСЕ, например, для ЛСЕ со значением номинации лица и группы лиц выделены следующие лексические классы: «Лицо», «Группа лиц», «Лицо и группа лиц».

Таблица ***Thematic_Groups_List*** содержит список тематических групп, по которым распределены ЛСЕ с отрицательно-оценочной семантикой (на данный момент – со значением номинации лица и группы лиц). Данный список сформирован с опорой на выделенные в РСС лексико-семантические множества, подмножества, группы, подгруппы и ряды.

Таблица ***Indecent_Form*** описывает возможность охарактеризовать форму негативной оценки (на данный момент – лица и группы лиц), выраженной при помощи ЛСЕ, как неприличную в трех вариантах: «можно», «вероятно», «нельзя», что определено на основании анализа зафиксированных в РСС СП при ЛСЕ.

Семантико-стилистическая характеристика отрицательно-оценочных слов в словарях дается в основных таблицах ЛБД ООЛ, формируемых отдельно для каждого словаря-источника, (например, в таблице ***Semantic_Stylistic_RSS*** для РСС). В основных таблицах ЛБД ООЛ каждая запись соответствует ЛСЕ с набором ее параметров, а каждое поле – единице информации об этой ЛСЕ, ее параметру (например, толкованию, типу значения, лексическому классу).

Каждая из таких таблиц имеет общую структуру и содержит следующие поля: *No* – порядковый номер записи в таблице; *Lexical_Semantic_Unit* – отрицательно-оценочная ЛСЕ; *Stylistic_Label* – отрицательно-оценочная СП при ЛСЕ (выбирается, например, из таблицы *Stylistic_Labels_RSS* для РСС); *Chrono_Label* – хронологическая помета при ЛСЕ (выбирается из таблицы *Chrono_Labels_List*); *Mean_Type* – тип значения ЛСЕ (выбирается из таблицы *Mean_Types_List*); *Definition* – толкование ЛСЕ; *Link* – отсылка к семантически близкой, тождественной, равнозначной ЛСЕ (выбирается, например, из таблицы *Semantic_Stylistic_RSS* для РСС); *Comment* – комментарий к толкованию ЛСЕ; *Lexical_Class* – лексический класс, к которому отнесена ЛСЕ (выбирается из таблицы *Lexical_Classes_List*); *Thematic_Group* – тематическая группа, в которую включена ЛСЕ (выбирается из таблицы *Thematic_Groups_List*); *Indecent_Form* – информация о возможности или невозможности охарактеризовать форму негативной оценки, выраженной при помощи ЛСЕ, как неприличную (выбирается из таблицы *Indecent_Form*); *Page_Number* – номер страницы словаря-источника, на которой расположена ЛСЕ.

Отметим, что в поле *Stylistic_Label* может быть выбрано несколько условных обозначений СП (с учетом их возможных вариантов, комбинаций, сочетаний, представленных в словаре), что связано с особенностями стилистической характеристики слов в словарях, когда при одном слове может насчитываться до пяти отрицательно-оценочных СП (подробнее об этом см. в [1]); а также в поле *Link* может быть внесено более одной семантически близкой или тождественной ЛСЕ, что обусловлено перечислением нескольких синонимов при синонимическом толковании слов в словарях-источниках.

***

Таким образом, рассмотрена структура ЛБД ООЛ и описаны связи между формирующими ее таблицами, в частности, для *Semantic_Stylistic_RSS* как одной из основных таблиц семантико-стилистической характеристики слов в РСС и *Stylistic_Labels_RSS*, *Chrono_Labels_List*, *Mean_Types_List*, *Lexical_Classes_List*, *Thematic_Groups_List*, *Indecent_ Form*, предназначенных для ее наполнения, а также ряда других вспомогательных таблиц.

3. Наполнение лингвистической базы данных отрицательно-оценочной лексики.

При заполнении основных таблиц данных ЛБД ООЛ – таблиц семантико-стилистической характеристики отрицательно-оценочных слов в словарях – из словарных статей лексикографических источников извлекаются следующие данные: отрицательно-оценочная ЛСЕ (на данный момент – со значением номинации лица и группы лиц); отрицательно-оценочная СП к ней при наличии; хронологическая помета к ней при наличии; семантическая помета к ней в случае переносного значения ЛСЕ; толкование значения ЛСЕ; в случае отсылочного и синонимического толкования из него выделяется семантически близкая, тождественная, равнозначная ЛСЕ; комментарий к толкованию при наличии.

Помимо собственно словарных данных в ЛБД ООЛ вносится информация о лексическом классе, к которому относится ЛСЕ; тематической группе, в которую входит ЛСЕ; возможности охарактеризовать как неприличную форму негативной оценки, выраженной при помощи ЛСЕ; номере страницы лексико-

графического издания, на которой располагается словарная статья, посвященная ЛСЕ.

Опишем наполнение данными одной из основных таблиц ЛБД ООЛ на примере первого тома РСС.

Таблица ***Semantic_Stylistic_RSS*** содержит 1 737 записей, соответствующих 1 737 отобранным из РСС отрицательно-оценочным ЛСЕ со значением номинации лица и группы лиц и их параметрам, и имеет общую для всех основных таблиц ЛБД ООЛ структуру, которая представлена выше. Охарактеризуем наполнение каждого поля данной таблицы.

В поле ***No*** указаны уникальные коды, являющиеся порядковыми номерами записей, а следовательно, и ЛСЕ, под которыми они внесены в данную таблицу с соблюдением идеографического принципа РСС: ЛСЕ сгруппированы по семантике и даны в алфавитном порядке в пределах каждой лексико-семантической группы.

Код является идентификатором записи в пределах данной таблицы, следовательно, отдельные отрицательно-оценочные значения одного многозначного слова зафиксированы как самостоятельные ЛСЕ под разными кодами в следующих случаях:

– отрицательно-оценочные значения многозначного слова даны в разных словарных статьях и отнесены к разным тематическим группам в РСС, например, для многозначного слова *Быдло* создано две записи: 1 765 *Быдло* (презр.) ‘люди, бессловесно выполняющие для кого-н. тяжелую работу’ и 1 829 *Быдло* (бран.) ‘о людях из низших социальных слоев’[2];

– отрицательно-оценочные значения многозначного слова приведены в одной словарной статье в пределах одной тематической группы в РСС, например, для многозначного слова *Сборище* создано две записи: 1 749 *Сборище* (неодобр.) ‘беспорядочное скопление людей’ и 1 750 *Сборище* ‘собрание людей (часто нелегальное)’;

– в отрицательно-оценочном значении можно выделить две части, на что указывает употребление в толковании сочетания союзов «а также», например, для слова *Кикимора* (также бран.) ‘человек, имеющий смешной и нелепый вид, а также некрасивая, непривлекательная женщина’ создано две записи: 1 545 *Кикимора* (также бран.) ‘человек, имеющий смешной и нелепый вид’ и 1 546 *Кикимора* (также бран.) ‘некрасивая, непривлекательная женщина’.

Кроме того, в ЛБД ООЛ как самостоятельные ЛСЕ зафиксированы:

– иллюстративные речения, которые получают стилистическую характеристику, отличную от характеристики заглавного слова, а также получают собственное толкование или развивают дополнительный (переносный) смысловой оттенок [6. С. XX], например, созданы такие записи: 1 812 *Беззащитное существо* (также ирон.) ‘живая особь, человек или животное’, 1 316 *Бог наказал* ‘о том, у кого случилась неприятность, кому не повезло, как будто бы в наказание за какую-то его вину, провинность’, 1 563 *Падший ангел* (также перен.) ‘о грешнике’;

– приведенные в так называемой заромбовой части словарной статьи устойчивые сочетания (фразеологические выражения и идиомы), «требующие собственного толкования и – в случае стилистического расхождения с заглавным словом – собственной стилистической характеристики» [6. С. XX], например, созданы такие записи: 7 *Человек в футляре* ‘человек, который замкнулся в кругу своих узких интересов, боится всего нового (по названию рассказа А. Чехова)’ и 483 *Лиса Патрикеевна* (пренебр.) ‘то же, что *Лиса*’ (см. 482 *Лиса* ‘хитрый и льстивый человек’);

– «ближайшие производные слова, мотивированные данным словозначением» [6. С. XX], в случае их отличия от производящего слова в стилистической характеристике, например, для слова *Жучок*, производного от слова *Жук* (см. 467 *Жук* (неодобр.) ‘ловкий человек, плут’), создана запись 468 *Жучок* (уменьш.-унич.) ‘то же, что *Жук*’, при этом производное слово представлено как новая ЛСЕ, для которой в поле *Definition* приведена идентифицирующая дефиниция «то же, что», а в поле *Link* указано производящее слово.

Также в ЛБД ООЛ созданы отдельные записи в случае наличия в зоне заглавного слова двух вариантов одной ЛСЕ, когда:

– оба варианта связаны союзом «и», например, для *Неумеха и Неумёха* создано две записи: 86 *Неумеха* (неодобр.) ‘человек, к-рый ничего не умеет делать или делает все плохо’ и 87 *Неумёха* ‘то же, что *Неумеха*’ – при этом второй вариант (после союза «и») представлен как новая ЛСЕ, для которой в поле *Definition* приведена идентифицирующая дефиниция «то же, что», а в поле *Link* указан первый вариант (перед союзом «и»);

– второй вариант присоединен при помощи союза «или», например, для *Фома неверный или неверующий* создано две записи: 209 *Фома неверный* ‘человек, к-рый упорно стоит на своем, не верит очевидному (по евангельской притче об апостоле Фоме, к-рый не хотел верить в воскресение Иисуса Христа до тех пор, пока не прикоснется к его ранам)’ и 210 *Фома неверующий* ‘то же, что *Фома неверный*’ – при этом второй вариант (после союза «или») представлен как новая ЛСЕ, для которой в поле *Definition* приведена идентифицирующая дефиниция «то же, что», а в поле *Link* указан первый вариант (перед союзом «или»);

– второй вариант помещен в скобки, например, для *Красна (красная) девица* создано две записи: 128 *Красна девица* (ирон.) ‘о чересчур скромном, застенчивом мужчине’ и 129 *Красная девица* ‘то же, что *Красна девица*’ – при этом второй вариант (в скобках) так же представлен как новая ЛСЕ, для которой в поле *Definition* так же приведена идентифицирующая дефиниция «то же, что», а в поле *Link* так же указан первый вариант (без скобок).

Наполнение поля ***Lexical_Semantic_Unit*** составляет 1 737 ЛСЕ, отобранных из РСС. В выборку вошли имена существительные и эквивалентные им единицы со значением номинации лица и группы лиц с отрицательным компонентом в денотативном значении, например, 1 292 *Противник* ‘враг, недруг’, 1 220 *Доносчик* ‘человек, к-рый доносит на кого-н., тайно обвиняет кого-н. перед властями, начальством’, – и с отрицательно-оценочным коннотативным компонентом значения, зафиксированным как в толковании, например, 1 307 *Бракодел* ‘недобросовестный работник, допускающий брак в своей работе’, 2 *Бумажная*

душа ‘о бездушном чиновнике, бюрократе’, – так и с помощью СП, например, 1354 *Аппаратчик* (часто неодобр.) ‘служащий какого-н. управления, руководящей организации, аппарата’, 1301 *Подельник* (неодобр.) ‘человек, участвующий в одном, общем с кем-н. деле’.

В общей сложности из РСС в ЛБД ООЛ (а именно в таблицу *Semantic_Stylistic_RSS*) включено 1 676 имен существительных (что составляет 96,5% от общего числа включенных единиц), а также 61 устойчивое сочетание (фразеологические выражения и идиомы), 4 пословицы и 2 поговорки (всего 3,5%).

При наполнении поля ***Stylistic_Label*** используются условные обозначения, передающие названия отрицательно-оценочных СП из системы помет, принятой в РСС, список которых представлен во вспомогательной таблице *Stylistic_Labels_RSS*. Данное поле заполнено для ЛСЕ, обладающих стилистической характеристикой, зафиксированной в РСС при помощи отрицательно-оценочных СП, при этом учтено их количество при ЛСЕ, комбинации СП, сочетания СП с кванторами. Так, в РСС отмечается наличие двух отрицательно-оценочных СП при ЛСЕ, встречаются комбинации СП *груб. бран.* и *презр. бран.*, возможны сочетания СП с кванторами. Отметим, что если в РСС отрицательно-оценочные СП при ЛСЕ отсутствуют, то данное поле остается незаполненным (ему присваивается значение пустой строки).

В таблице *Semantic_Stylistic_RSS* насчитывается 720 ЛСЕ (что составляет 41,5% от их общего числа), обладающих стилистической характеристикой, зафиксированной при помощи отрицательно-оценочных СП: из них с СП *ирон*. – 86, *неодобр*. – 142, *пренебр*. – 163, *унич*. – 13, *презр*. – 160, *груб*. – 35, *бран*. – 197, а также с комбинациями СП *груб. бран*. – 18, *презр. бран*. – 2. Отметим, что из числа стилистически маркированных при 104 ЛСЕ (14,4%) имеется две отрицательно-оценочные СП, что свидетельствует об их полифункциональном употреблении.

Поле ***Chrono_Label*** данной таблицы заполняется информацией о выходящих (при помощи хронологической пометы *устар*.) и вышедших (при помощи хронологической пометы *стар*.) из употребления ЛСЕ, что дополняет их основную стилистическую (отрицательно-оценочную) характеристику. В данном поле для ЛСЕ указана их временна́я характеристика (при наличии таковой в РСС), для чего использованы условные обозначения хронологических помет из списка, представленного во вспомогательной таблице *Chrono_Labels_List*. Отметим, что если ЛСЕ входят в активный запас, то данное поле остается незаполненным (ему присваивается значение пустой строки).

В таблице *Semantic_Stylistic_RSS* насчитывается 1 622 ЛСЕ (что составляет 93,4% от их общего числа), находящихся в активном запасе; 108 ЛСЕ (6,2%), выходящих из употребления, например, 340 *Мироед* (презр.; устар.) ‘Человек, живущий чужим трудом, паразит’; 7 ЛСЕ (0,4%), вышедших из употребления, например, 682 *Лиходей* (также бран.; стар.) ‘то же, что *Злодей*’.

В поле ***Mean_Type*** делается отметка в виде семантической пометы *перен*. (выбирается из списка типов значения, представленного во вспомогательной таблице *Mean_Types_List*) в случае зафиксированного в РСС переносного значения ЛСЕ, что уточняет ее основную семантическую характеристику – значение номинации лица и группы лиц. Отметим, что если значение ЛСЕ прямое, то данное поле остается незаполненным (ему присваивается значение пустой строки).

В таблице *Semantic_Stylistic_RSS* насчитывается 1 500 ЛСЕ (что составляет 86,4% от их общего числа) в прямом значении и 237 ЛСЕ (13,6%) в переносном значении, например, 40 *Бревно* (пренебр., также бран.; перен.) ‘Глупый и грубый человек’, 52 *Колпак* (пренебр., также бран.; перен.) ‘Глупый, недалекий, простоватый человек’, 103 *Щенок* (пренебр., бран.; перен.) ‘Мальчишка, молокосос’.

В поле ***Definition*** приводятся толкования ЛСЕ из РСС. Для многозначных слов, зафиксированных в РСС в одной или разных тематических группах, отбирается только то значение (реже – несколько значений), которое выражает семантику отрицательной оценки, и фиксируется в качестве толкования отрицательно-оценочной ЛСЕ. Так, содержащиеся в данной таблице 1 737 ЛСЕ представляют собой 1 612 многозначных слов в их конкретных значениях.

Помимо собственно семантического толкования в РСС используется два приема: синонимическое толкование и отсылочное (опосредованное). Опишем особенности заполнения поля *Definition* в данных случаях.

Приведенное в данном поле синонимическое толкование для ЛСЕ может быть как единственным для нее, например, 1 548 *Страхолюдина* ‘урод, страшилище’; так и дополнением к семантическому толкованию, например, 352 *Сачок* ‘человек, отлынивающий от работы, бездельник, лентяй’. Отметим, что наряду с заполнением данного поля перечисленные в толковании синонимичные ЛСЕ дублируются в поле *Link* (при условии их наличия в данной таблице).

В таблице *Semantic_Stylistic_RSS* 439 ЛСЕ (что составляет 25,3% от их общего числа) толкуется через другие семантически близкие или тождественные им единицы.

Отсылочное (опосредованное) толкование в РСС семантически равнозначных ЛСЕ с помощью идентифицирующей дефиниции «то же, что» фиксируется аналогичным образом: в данном поле приводится указанная дефиниция с внесением в поле *Link* той ЛСЕ, отсылка к которой дана в толковании. В данной таблице при отсылочном (опосредованном) толковании встречаются отсылки также к тем ЛСЕ, которые в ней не представлены, например, *Бабка* (иногда пренебр.) ‘то же, что *Бабушка*’, *Студентик* (пренебр.) ‘то же, что *Студент*’. В приведенных примерах определяемые ЛСЕ являются отрицательно-оценочными, а определяющие – нет (следовательно, не включены в ЛБД ООЛ), однако так как определяемые ЛСЕ в РСС не имеют собственного толкования кроме идентификации «то же, что», то толкования для таких отрицательно-оценочных ЛСЕ, как *Бабка*, *Студентик* приводятся от соответствующих нейтральных ЛСЕ *Бабушка*, *Студент*, а в поле *Link* факт отсылки к ним не

указывается, например, 1 248 *Бабка* (иногда пренебр.) ‘мать отца или матери’, 1 339 *Студентик* (пренебр.) ‘учащийся высшего учебного заведения’.

В таблице *Semantic_Stylistic_RSS* 211 ЛСЕ (что составляет 12,2% от их общего числа) не имеют самостоятельного значения, а толкуются через отсылку к другим ЛСЕ, представленным в ней.

В поле ***Link*** фиксируются семантические связи между семантически близкими, тождественными, равнозначными ЛСЕ, представленными в ЛБД ООЛ, в случаях их синонимического и отсылочного (опосредованного) толкования в РСС (как представлено выше при описании поля *Definition*), а также для ближайших производных от заглавного слова и двух вариантов одной ЛСЕ, приведенных в РСС в зоне заглавного слова (как представлено выше при описании поля *No*). При заполнении данного поля для определяемой ЛСЕ выбирается определяющая ЛСЕ из данной таблицы (или более одной определяющей ЛСЕ, когда в РСС встречается перечисление в толковании к ЛСЕ нескольких синонимов). Отметим, что если такие связи между ЛСЕ отсутствуют, то данное поле остается незаполненным (ему присваивается значение пустой строки).

В таблице *Semantic_Stylistic_RSS* для 245 ЛСЕ (что составляет 14,1% от их общего числа) установлены семантические связи в виде отсылок к другим ЛСЕ.

Поле ***Comment*** содержит извлеченные из словарных статей РСС комментарии к толкованию ЛСЕ, представляющие собой дополнительную информацию, позволяющую уточнять отрицательно-оценоч-ную семантику ЛСЕ. В данном поле представлено несколько типов комментариев, различающихся по своей функции: этимологический комментарий, например, 1 345 *Эскулап* (ирон.) ‘врач, лекарь (по имени бога врачевания в древнеримской мифологии)’; приведение первоначального значения, например, 700 *Бурбон* (презр.; устар.) ‘грубый, невоспитанный человек, солдафон (первоначально об офицере, выслужившемся из низов)’; указание на локально-временну́ю характеристику, например, 996 *Вредитель* ‘человек, к-рому приписывается умышленное нанесение вреда производству (в годы советской власти)’; уточнение объекта отрицательной оценки (его пола, возраста, рода деятельности, других характеристик), например, 1087 *Марионетка* (перен.) ‘человек, покорно и слепо действующий по чужой воле (обычно о политическом деятеле)’; уточнение оттенка оценочного значения, например, 1 662 *Поросенок* (бран.) ‘употребляется как бранное слово (с оттенком насмешки)’; указание на сферу употребления, например, 1 721*Чернь* (презр.; устар.) ‘низшие слои общества, трудовой народ (в речи аристократов, власть имущих)’; демонстрация лексической сочетаемости, например, 1 811 *Оборотень* (неодобр.) ‘о хитром и двуличном человеке (в сочетании: Не человек, а прямо какой-то оборотень)’; указание на особенности функционирования в речи, например, 4 *Тип* (ирон.) ‘человек, отличающийся какими-н. характерными свойствами, приметами (обычно с определением)’; определение лексического статуса единицы, например, 1 467 *Горбатого могила исправит* ‘о том, чьи недостатки неисправимы (пословица)’; уточнение морфологических характеристик, например, 1 707 *Низ* ‘непривилегированные слои общества (во мн. ч.)’. Отметим, что если комментарий к толкованию ЛСЕ в словарной статье РСС отсутствует, то данное поле остается незаполненным (ему присваивается значение пустой строки).

В таблице *Semantic_Stylistic_RSS* 173 ЛСЕ (что составляет 10% от их общего числа) снабжено комментариями.

В поле ***Lexical_Class*** указываются названия лексических классов (выбираются из списка, представленного во вспомогательной таблице *Lexical_ Classes_List*). Основанием для отнесения ЛСЕ к одному из них является в первую очередь семантика ЛСЕ (ее включенность в РСС в соответствующее лексическое множество), а также наличие семантической пометы *собир.*, свидетельствующей о принадлежности ЛСЕ к разряду собирательных имен существительных, например, 1787 *Пьянь* (презр.) ‘Пьяные, пьющие люди’. Отметим, что к классу «Лицо и группа лиц» отнесены ЛСЕ с пометой *собир.* в сочетании с квантором *также*, поскольку такие ЛСЕ могут обозначать как отдельное лицо, так и совокупность лиц, например, 65 *Полудурье* ‘то же, что *Полудурок*’.

В таблице *Semantic_Stylistic_RSS* представлено 1 621 ЛСЕ (что составляет 93,3% от их общего числа), обозначающих лицо; 109 ЛСЕ (6,3%), обозначающих группу лиц; 7 ЛСЕ (0,4%), обозначающих лицо и группу лиц.

При заполнении поля ***Thematic_Group*** для каждой ЛСЕ выбирается название тематической группы (из списка, представленного во вспомогательной таблице *Thematic_Groups_List*), соответствующее тому лексическому множеству, к которому ЛСЕ отнесена в РСС.

В поле ***Indecent_Form*** фиксируется возможность охарактеризовать форму негативной оценки, выраженной при помощи ЛСЕ, как неприличную (с использованием условных обозначений из вспомогательной таблицы *Indecent_Form*), которая определяется на основании анализа СП при ЛСЕ, представленных в поле *Stylistic_Label* данной таблицы. Так, условное обозначение «можно» выбирается для ЛСЕ, снабженных СП *бран.*, *груб.*, а также комбинациями СП *груб. бран., груб. презр.*; «вероятно» – для ЛСЕ с теми же СП в сочетании с кванторами; «нельзя» – для ЛСЕ с другими отрицательно-оценочными СП или при отсутствии таковых.

В таблице *Semantic_Stylistic_RSS* насчитывается 114 ЛСЕ (что составляет 6,6% от их общего числа), для которых, по данным РСС, форму выражаемой ими негативной оценки можно охарактеризовать как неприличную; 99 ЛСЕ (5,7%), для которых, по данным РСС, форму выражаемой ими негативной оценки, вероятно, можно охарактеризовать как неприличную; 1 524 ЛСЕ (87,7%), для которых, по данным РСС, форму выражаемой ими негативной оценки охарактеризовать как неприличную нельзя.

Поле ***Page_Number*** содержит информацию о постраничном расположении в словнике РСС словарных статей, посвященных ЛСЕ. При этом если словарная статья начинается на одной странице, а заканчивается на другой, то в данном поле записываются оба номера страницы через дефис. Данная информация необхо-

дима как для навигации по словарю-источнику при работе с ЛБД ООЛ, так и для формирования ссылки на него с подстановкой номера страницы в библиографическое описание, хранящееся во вспомогательной таблице *Dictionaries_List*.

***

Таким образом, на примере одного из лексикографических источников рассмотрено строение и наполнение основной таблицы ЛБД ООЛ – таблицы семантико-стилистической характеристики отрицательно-оценочных слов в РСС, структуру которой формируют 12 полей, содержащих информацию как лингвистического, так и вспомогательного характера, что позволяет предоставлять релевантную информацию об отрицательно-оценочных ЛСЕ как из данного словаря, так и в сопоставлении с другими источниками.

Так, при использовании ЛБД ООЛ в качестве исследовательского инструмента: содержащиеся в поле *No* уникальные коды ЛСЕ позволяют различать значения многозначных слов; наполнение поля *Lexical_Semantic_Unit* дает представление о составе пласта отрицательно-оценочной лексики в современном русском языке по данным словаря; зафиксированные в поле *Stylistic_Label* отрицательно-оценочные СП отражают стилистическую маркированность ЛСЕ в словаре; содержащиеся в поле *Chrono_Label* хронологические пометы позволяют отбирать лексику, относящуюся к активному запасу; указание в поле *Mean_Type* на переносный тип значения позволяет определять в контексте слова, развивающие дополнительный смысловой оттенок; поле *Definition* предоставляет толкования ЛСЕ; реализованные при помощи поля *Link* связи между ЛСЕ позволяют прослеживать системные отношения между единицами лексического множества отрицательно-оценочных имен существительных со значением номинации лица и группы лиц; комментарии из поля *Comment* дополняют семантику ЛСЕ; поля *Lexical_Class* и *Thematic_Group* отражают распределение ЛСЕ по соответствующим лексическим классам и тематическим группам; поле *Indecent_Form* позволяет определять форму выраженной при помощи ЛСЕ отрицательной оценки как неприличную на основании словарных данных; зафиксированные в поле *Page_ Number* номера страниц лексикографического издания выполняют функции навигации по нему и ссылки на полученную с помощью ЛБД ООЛ информацию.

Выводы

В результате проведенного исследования отобран, зафиксирован, систематизирован, семантически и стилистически описан пласт отрицательно-оценочных ЛСЕ и разработана концептуальная структура описания таких единиц, получившая отражение в формате ЛБД.

ЛБД ООЛ предоставляет информацию как лингвистического, так и вспомогательного характера об отрицательно-оценочных ЛСЕ, позволяет осуществлять их поиск, сортировку и отбор по выделенным критериям (параметрам), а также сопоставлять информацию из нескольких источников.

Основная область применения ЛБД ООЛ – лингвоэкспертная практика. Также данная ЛБД открывает новые перспективы для дальнейших разработок в области теоретической и практической лексикографии, выступая источником для составления словарей разных типов (например, словаря отрицательно-оценочной лексики, словаря наименований лиц, словаря инвективной лексики).

Формат представления лексикографических данных в виде ЛБД позволяет реализовать новый подход к проведению лингвистических исследований – с использованием средств автоматизации (например, для автоматического поиска оценочной лексики в конфликтных текстах), а также проводить экспериментальные исследования по определению семантики и стилистической характеристики слов (в том числе подбирать материал для проведения психолингвистического эксперимента, обрабатывать полученные данные), тем самым пополняя ЛБД ООЛ их результатами.

В качестве перспективы пополнения ЛБД ООЛ новыми данными отметим также возможность привлечения «Национального корпуса русского языка» в качестве источника материала для уточнения значений и стилистической квалификации слов, а также включение в ЛБД ООЛ единиц других частей речи и других семантических и стилистических групп лексики.

Подобная ЛБД может быть не только использована в качестве исследовательского инструмента, но и применена в образовательном процессе для ряда дисциплин (например, «Прикладная лингвистика», «Юридическая лингвистика», «Лексикография», «Лексикология»), а также в других видах деятельности (например, в изучении и преподавании русского языка как родного, неродного, иностранного и переводческой деятельности).

ПРИМЕЧАНИЕ

[1] Здесь и далее в приведенных примерах подчеркиванием выделены те фрагменты записей ЛБД ООЛ, на которых акцентируется внимание при анализе.

ЛИТЕРАТУРА

1. Кочергина К.С. Стилистические пометы в толковых словарях современного русского языка: сопоставительный анализ // Вопросы лексикографии. 2017. № 11. С. 20–38.
2. Голев Н.Д. Юрислингвистический словарь инвективной лексики русского языка (к постановке проблемы) // Актуальные проблемы русистики : материалы Международной конференции / отв. ред. Т.А. Демешкина. Томск : Изд-во Том. ун-та, 2003. Вып. 2, Ч. 1. С. 92–98.
3. Кусов Г.В. Выработка рекомендаций по совершенствованию информационных технологий в практике производства судебной лингвистической экспертизы // Теория и практика общественного развития. 2012. № 10. С. 321–326.

4. Кочергина К.С. Система интеллектуальной поддержки процедуры лингвистической экспертизы // Перспективы развития фундаментальных наук : тр. XII Междунар. конф. студ. и молодых ученых (Томск, 21–24 апреля 2015 г.). Томск : Изд-во Том. политехн. ун-та, 2015. С. 1497–1499.
5. База данных // Большой энциклопедический словарь / под ред. А.М. Прохорова. М. : Сов. энциклопедия, 1993. С. 96.
6. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / под общей ред. Н.Ю. Шведовой. М. : Азбуковник, 1998. Т. 1 : Слова указующие (местоимения). Слова именующие: имена существительные (Все живое. Земля. Космос). 800 с.
7. Мишанкина Н.А. Базы данных в лингвистических исследованиях // Вопросы лексикографии. 2013. № 1 (3). С. 25–33.
8. Мишанкина Н.А. Базы данных в филологических исследованиях // Открытое и дистанционное образование. 2012. Т. 2. С. 35–40.
9. Мишанкина Н.А., Тубалова И.В., Эмер Ю.А. Филология и информатика : специфика электронного представления региональных фольклорных текстов // Гуманитарная информатика : сб. ст. / под ред. Г.В. Можаевой. Томск: Изд-во Том. ун-та, 2004. Вып. I. С. 102–114.
10. Баранов А.Н. Введение в прикладную лингвистику : учеб. пособие. М. : Эдиториал УРСС, 2001. 360 с.
11. Кочергина К.С. Применение технологии базы данных в лингвистических исследованиях // Российская академическая лексикография: современное состояние и перспективы развития : сб. науч. ст. по материалам Междунар. науч. конф., посвященной 70-летию выхода первого тома академического «Словаря современного русского литературного языка» / отв. ред. О.Н. Крылова, С.А. Мызников, М.Н. Приемышева, Е.В. Пурицкая. СПб. : Нестор-История, 2018. С. 158–169.
12. Мишанкина Н.А., Панасенко Е.А. База данных метафорической терминологии : концептуальное проектирование // Вестник Новосибирского государственного педагогического университета. 2016. № 6 (34). С. 86–99.
13. Резанова З.И., Миклашевский А.А. Моделирование образно-перцептивного компонента языковой семантики при помощи психолингвистической базы данных // Вестник Томского государственного университета. Филология. 2016. № 5 (43). С. 71–92.
14. Булыгина Е.Ю., Трипольская Т.А. Национально-культурный компонент в семантике слова и способы его представления в базе данных прагматически маркированной лексики // Вопросы лексикографии. 2017. № 11. С. 5–19.
15. Булыгина Е.Ю., Трипольская Т.А. База данных прагматически маркированной лексики русского языка : материал, принципы описания, возможности использования // Вестник Новосибирского государственного педагогического университета. 2016. № 6 (34). С. 70–85.
16. Басалаева Е.Г. Прагматический макрокомпонент и способы его семантизации в электронной базе данных // Вестник Новосибирского государственного педагогического университета. 2016. № 6 (34). С. 112–125.
17. Саженин И.И. К вопросу о построении базы данных прагматически маркированной лексики // Вестник Новосибирского государственного педагогического университета. 2015. № 5 (27). С. 114–121.
18. Пурицкая Е.В., Панков Д.И. Нормативно-стилистическая характеристика лексики современного русского языка: возможности описания в словарной базе данных // Вопросы лексикографии. 2018. № 13. С. 23–43.
19. Свидетельство о государственной регистрации базы данных № 2018620042. Лингвистическая база данных отрицательно-оценочной лексики (имена существительные со значением номинации лица и группы лиц) / К.С. Кочергина. М. : Роспатент, 2018.
20. Кочергина К.С. Проектирование базы лингвистических данных (на материале оценочной лексики) // Материалы XIX Открытой конференции студентов-филологов (Санкт-Петербург, 18–22 апреля 2016 г.) / отв. ред. С.И. Монахов. СПб. : Изд-во СПбГУ, 2017. С. 170–174.
21. Матвеева О.Н. Функционирование конфликтных текстов в правовой сфере и особенности его лингвистического изучения (на материале текстов, вовлеченных в юридическую практику) : дис. ... канд. филол. наук. Барнаул, 2004. 283 с.
22. Гибатова Г.Ф. Аксиология в языке // Вестник Оренбургского государственного университета. 2011. № 2 (121). С. 127–132.
23. Иваненко Г.С. Лексикография и лингвистическая экспертиза : перспективы взаимоотношений // Юрислингвистика. 2018. № 7–8. С. 98–118.
24. Голев Н.Д., Матвеева О.Н. Юрислингвистическая экспертиза : на стыке языка и права // Сибирский филологический журнал. 2003. № 1. С. 146–157.
25. Ефремов В.А. Словари субстандарта в лингвоэкспертной практике // Вестник Новосибирского государственного педагогического университета. 2015. № 2 (24). С. 41–49.
26. Стернин И.А. Основные принципы семантического анализа в лингвокриминалистической экспертизе текста // Вестник Кемеровского государственного университета. 2017. № 1 (69). С. 202–207.
27. Кочергина К.С. Автоматизация маркирования оценочной лексики в экспертных текстах // XVIII Международная конференция студентов-филологов СПбГУ (Санкт-Петербург, 6–11 апреля 2015 г.) : тез. докл. / отв. ред. Д.Н. Чердаков. СПб. : Филологический факультет СПбГУ, 2015. С. 247–248.

Статья представлена научной редакцией «Филология» 30 июля 2019 г.

Linguistic Database of Negative Evaluative Lexis: Concept, Structure, Content

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2019, 446, 29–40.
DOI: 10.17223/15617793/446/4
Kristina S. Kochergina, Tomsk State University (Tomsk, Russian Federation). E-mail: tinakochergina@gmail.com
Keywords: linguistic database; noun; negative evaluative lexis; stylistic label; dictionary; juridical linguistics; linguistic expert examination.

The article is devoted to the Linguistic Database of Negative Evaluative Lexis (Nouns Denoting Person and Group of Persons)" (LDoNEL) which is being developed for linguistic research, in particular, for linguistic expert examination. The aim of this article is to describe LDoNEL's concept, structure and content. The sources for LDoNEL are the dictionaries of the modern Russian language (from late 20th to early 21st centuries). The material for the study were nouns with negative evaluative semantics denoting a person and a group of persons. The nouns were selected from the indicated sources by continuous sampling based on ideographic, semantic (including thematic one) and stylistic principles. A descriptive analytical (general scientific) method with a combination of its techniques was used in the study. The conceptual design method was used in developing LDoNEL. LDoNEL is characterized as follows: according to the type of lexical units, namely, by their part-of-speech affiliation, it contains nouns; according to the lexical semantic description it describes the negative-evaluative component; according to the data source it contains data from dictionaries and information based on them (about thematic groups and indecent forms); according to application it is intended for use in forensic linguistics practice and also can be used for various research and practical purposes. The article describes the structure of LDoNEL, which consist of 11 data tables: the main table of semantic and stylistic characteristics of negative evaluative lexis in the dictionary, and 10 supporting data tables ensuring the coherence and integrity of the data in the main tables: a dictionaries table; a negative evaluative lexical units table; a lexical units types table; a negative evaluative stylistic labels table; a dictionary stylistic labels table; a chronological labels table; a lexical meaning types table; a lexical classes table; a thematic groups table; a table of characteristics of the form of negative evaluation as indecent. Table fields that form the table's structure are listed for each of the tables. The content of LDo-

NEL is presented on the example of the first volume of the *Russian Semantic Dictionary* (1998) edited by N.Yu. Shvedova which is the basis for the main table of semantic and stylistic characteristics of negative evaluative lexis in the dictionary. The main table contains 1,737 entries relevant to 1,737 negative evaluative nouns denoting a person and a group of persons and their parameters. As a result of the study, a layer of negative evaluative lexis was selected, fixed, systematized, semantically and stylistically described, and a conceptual structure for describing such units was developed and reflected in the LDoNEL format.

REFERENCES

1. Kochergina, K.S. (2017) Stylistic Labels in Dictionaries of the Modern Russian Language: a Comparative Analysis. *Voprosy leksikografii – Russian Journal of Lexicography*. 11. pp. 20–38. (In Russian). DOI: 10.17223/22274200/11/2
2. Golev, N.D. (2003) [Legal Linguistics Dictionary of the Invective Vocabulary of the Russian Language (towards the Problem)]. *Aktual'nye problemy rusistiki* [Actual Problems of Russian Studies]. Proceedings of the International Conference. 2 (1). Tomsk. 21–23 October 2003. Tomsk: Tomsk State University. pp. 92–98. (In Russian).
3. Kusov, G.V. (2012) Development of Guidelines for the Improvement of Information Technologies Implementation in Forensic Linguistic Expertise. *Teoriya i praktika obshchestvennogo razvitiya – Theory and Practice of Social Developement*. 10. pp. 321–326. (In Russian).
4. Kochergina, K.S. (2015) [The Intelligent Support System for a Linguistic Expertise]. *Perspektivy razvitiya fundamental'nykh nauk* [Prospects of Fundamental Sciences Development]. Proceedings of the 12th International Conference. Tomsk. 21–24 April 2015. Tomsk: Tomsk Polytechnic University. pp. 1497–1499. (In Russian).
5. Prohorov, A.M. (ed.) (1993) Baza dannykh [Database]. In: *Bol'shoy entsiklopedicheskiy slovar*' [Large Encyclopedic Dictionary]. Moscow: Sov. entsiklopediya.
6. Shvedova, N.Yu. (ed.) (1998) *Russkiy semanticheskiy slovar*'. *Tolkovyy slovar*', *sistematizirovannyy po klassam slov i znacheniy* [Russian Semantic Dictionary. Explanatory Dictionary, Systematized by the Classes of Words and Meanings]. Vol. 1. Moscow: Azbukovnik.
7. Mishankina, N.A. (2013) Databases in Linguistic Research. *Voprosy leksikografii – Russian Journal of Lexicigraphy*. 1 (3). pp. 25–33. (In Russian).
8. Mishankina, N.A. (2012) Database In Socio-Humanitarian Research. *Otkrytoe i distantsionnoe obrazovanie – Open and Distance Education*. 2 (46). pp. 35–40.
9. Mishankina, N.A., Tubalova, I.V. & Emer, Yu.A. (2004) Filologiya i informatika: spetsifika elektronnogo predstavleniya regional'nykh fol'klornykh tekstov [Philology and Computer Science: the Specifics of the Electronic Representation of Regional Folklore Texts]. In: Mozhaeva, G.V. (ed.) *Gumanitarnaya informatika* [Humanitarian Computer Science]. 1. Tomsk: Tomsk State University. pp. 102–114.
10. Baranov, A.N. (2001) *Vvedenie v prikladnuyu lingvistiku* [Introduction to Applied Linguistics]. Moscow: Editorial URSS.
11. Kochergina, K.S. (2018) [The Use of Database Technology in Linguistic Research]. *Rossiyskaya akademicheskaya leksikografiya: sovremennoe sostoyanie i perspektivy razvitiya* [Russian Academic Lexicography: Current Status and Development Prospects]. Proceedings of the International Conference dedicated to the 70th anniversary of the publication of the first volume of the academic "Dictionary of the Modern Russian Literary Language". St. Petersburg. 3–5 October 2018. St. Petersburg: Nestor-Istoriya. pp. 158–169. (In Russian).
12. Mishankina, N.A. & Panasenko, E.A. (2016) The Metaphorical Terminology Database: Conceptual Design. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta – Novosibirsk State Pedagogical University Bulletin*. 6 (34). pp. 86–99. (In Russian). DOI: 10.15293/2226-3365.1606.07
13. Rezanova, Z.I. & Miklashevskiy, A.A. (2016) Modeling of the Perceptual-Based Component of Language Semantics Using a Psycholinguistic Database. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 5 (43). pp. 71–92. (In Russian). DOI: 10.17223/19986645/43/6
14. Bulygina, E.Yu. & Tripol'skaya, T.A. (2017) The National-Cultural Component in the Semantics of Words and Ways of Presenting It in a Database of Pragmatically Marked Vocabulary. *Voprosy leksikografii – Russian Journal of Lexicography*. 11. pp. 5–19. (In Russian). DOI: 10.17223/22274200/11/1
15. Bulygina, E.Yu. & Tripol'skaya, T.A. (2016) A Database of Pragmatically Marked Lexical Items of the Russian Language: The Content, Principles of Description, and Possibility of Using. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta – Novosibirsk State Pedagogical University Bulletin*. 6 (34). pp. 70–85. (In Russian). DOI: 10.15293/2226-3365.1606.06
16. Basalaeva, E.G. (2016) The Pragmatic Component and Ways of Its Semantization in the Electronic Database. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta – Novosibirsk State Pedagogical University Bulletin*. 6 (34). pp. 112–125. (In Russian). DOI: 10.15293/2226-3365.1606.09
17. Sazhenin, I.I. (2015) About Construction of a Database of Pragmatically Marked Vocabulary. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta – Novosibirsk State Pedagogical University Bulletin*. 5 (27). pp. 114–121. (In Russian). DOI: 10.15293/2226-3365.1505.10
18. Puritskaya, E.V. & Pankov, D.I. (2018) The Lexicon of the Modern Russian Language: Describing Normative and Stylistic Characteristics in the Dictionary Database. *Voprosy leksikografii – Russian Journal of Lexicography*. 13. pp. 23–43. (In Russian). DOI: 10.17223/22274200/13/2
19. Kochergina, K.S. (2018) *Svidetel'stvo o gosudarstvennoy registratsii bazy dannykh № 2018620042. Lingvisticheskaya baza dannykh otritsatel'no-otsenochnoy leksiki (imena sushchestvitel'nye so znacheniem nominatsii litsa i gruppy lits)* [Certificate of State Registration of a Database No. 2018620042. Linguistic Database of Negative Evaluative Lexics (Nouns Naming a Person or a Group of People)]. Moscow: Rospatent.
20. Kochergina, K.S. (2017) [Designing a Database of Linguistic Data (Based on Evaluative Vocabulary)]. *Proceedings of the 19th Open Conference for Philology Students*. St. Petersburg. 18–22 April 2016. St. Petersburg: St. Petersburg State University. pp. 170–174. (In Russian).
21. Matveeva, O.N. (2004) *Funktsionirovanie konfliktnykh tekstov v pravovoy sfere i osobennosti ego lingvisticheskogo izucheniya (na materiale tekstov, vovlechennykh v yuridicheskuyu praktiku)* [The Functioning of Conflicting Texts in the Legal Sphere and the Peculiarities of Its Linguistic Study (Based on the Texts of Texts Involved in Legal Practice)]. Philology Cand. Diss. Barnaul.
22. Gibatova, G.F. (2011) Axiology at the Language. *Vestnik Orenburgskogo gosudarstvennogo universiteta – Vestnik of the Orenburg State University*. 2 (121). pp. 127–132. (In Russian).
23. Ivanenko, G.S. (2018) Lexicography and Linguistic Expertise: Prospects of Mutual Relations. *Yurislingvistika – Legal Linguistics*. 7 (8). pp. 98–118. (In Russian).
24. Golev, N.D. & Matveeva, O.N. (2003) Yurislingvisticheskaya ekspertiza: na styke yazyka i prava [Legal Lingistic Examination: at the Intersection of Language and Law]. *Sibirskiy filologicheskiy zhurnal – Siberian Journal of Philology*. 1. pp. 146–157.
25. Efremov, V.A. (2015) Dictionaries of Substandart Language in Lingvo Expert Practice. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta – Novosibirsk State Pedagogical University Bulletin*. 2 (24). pp. 41–49. (In Russian). DOI: 10.15293/2226-3365.1502.03
26. Sternin, I.A. (2017) Main Principles of Semantic Analysis in Linguocriminalistic Text Examination. *Vestnik Kemerovskogo gosudarstvennogo universiteta – Bulletin of Kemerovo State University*. 1 (69). pp. 202–207. (In Russian). DOI: 10.21603/2078-8975-2017-1-202-207
27. Kochergina, K.S. (2015) [Automatisation of Marking of Evaluative Vocabulary in Expert Texts]. *Abstracts of the 18th Open Conference for Philology Students*. St. Petersburg. 6–11 April 2015. St. Petersburg: Saint Petersburg State University. pp. 247–248. (In Russian).

Received: 30 July 2019