

*Кочергина Кристина Сергеевна*

Томский государственный университет (ТГУ),  
Томск, Россия  
tinakohergina@gmail.com

## **ПРОЕКТИРОВАНИЕ БАЗЫ ЛИНГВИСТИЧЕСКИХ ДАННЫХ (НА МАТЕРИАЛЕ ОЦЕНОЧНОЙ ЛЕКСИКИ)**

**Ключевые слова:** база данных, оценочная лексика, конфликтный текст, лингвистическая экспертиза.

Статья посвящена описанию принципов формирования и структуры лингвистической базы данных отрицательно-оценочных лексико-семантических единиц. Охарактеризовано наполнение столбцов таблицы базы данных. В качестве основного источника наполнения базы данных использован «Русский семантический словарь» (первый том). Проектируемая база данных предназначена для автоматического поиска представленной в ней лексики при экспертировании конфликтных текстов.

*Kochergina Kristina*

Tomsk State University (TSU),  
Tomsk, Russia

## **DESIGNING THE LINGUISTIC DATABASE OF NEGATIVE EVALUATIVE LEXICON**

**Keywords:** database, evaluative lexicon, conflict text, linguistic expertise.

The article is devoted to the description of the principles of formation and the structure of the negative evaluative lexicon database. The fields of the database are described. As the main database content source was used "Russian Semantic Dictionary" (the first volume). The designed database is intended for automatic searching negative evaluative lexicon in the field of conflict texts expertise.

Одним из перспективных направлений на пересечении областей прикладной лингвистики и современной лексикографии является создание баз лингвистических данных (далее: БД) [Мишанкина, 2013]. Формат БД обладает рядом преимуществ перед традиционными толковыми словарями, поскольку позволяет хранить систематизированную информацию и осуществлять быстрый поиск, а также имеет открытую структуру для внесения изменений [Баранов, 2001].

Цель настоящей статьи заключается в описании принципов и структуры лингвистической БД, составленной на материале отрицательно-оценочных лексико-семантических единиц (далее: ЛСЕ). БД составлена в соответствии с четырьмя основополагающими принципами.

**Принцип первый — опора на авторитетные толковые словари.** Поскольку предмет данного исследования ограничен нормативными лексическими единицами литературного языка, то в качестве источников материала для БД целесообразно использовать толковые словари (к тому же только этот вид источников признаётся в экспертной практике легитимным [Голев, Матвеева, 2003]). Под авторитетными понимаются академические лексикографические издания, активно применяющиеся в экспертной практике. Основным словарём-источником для данной БД стал первый том «Русского семантического словаря» под ред. Н. Ю. Шведовой (далее: РСС) [Шведова, 1998].

**Принцип второй — отбор лексики.** В данной работе принимается во внимание нормативная оценочная лексика с отрицательным компонентом значения, что шире понятия «инвективная лексика». Инвективная лексика выполняет функцию оскорбления адресата, в то время как отрицательно-оценочная лексика может не только реализовать инвективное намерение, но и передать различного рода негативную информацию об объекте оценки. Такой подход позволяет определять наличие в конфликтном тексте негативной информации для решения вопросов по различным категориям дел: не только об оскорблении, но и клевете, призывах, угрозах и пр.

В состав отрицательно-оценочной лексики включена литературная (в том числе разговорная) лексика как с коннотативным оценочным значением, зафиксированным в словарях с помощью стилистических помет, так и денотативным оценочным значением: прямым и переносным.

В фокусе внимания находится ограниченная группа слов — только имена существительные с семантикой номинации лица и совокупности лиц. С точки зрения частеречной принадлежности, отбираются именно существительные, несмотря на то, что выражать негативную оценку могут и другие части речи. Это объясняется тем, что имена существительные содержат оценочную информацию о самом объекте оценки — лице (группе лиц) — в то время как, например, глаголы характеризуют не лицо, а его действия, а прилагательные обозначают характеристики лица.

**Принцип третий — способ представления данных.** БД представляет собой перечень ЛСЕ и их параметров, среди которых отсутствуют такие традиционные лексикографические атрибуты слов, как грамматические характеристики и контексты. Отсутствие необходимости указания грамматических сведений о каждой ЛСЕ обусловлено тем, что на этапе предобработки текста они определяются автоматически. Каждая ЛСЕ не иллюстрируется примерами употребления, поскольку эта БД предназначена для поиска ЛСЕ и их реальных контекстов употребления с возможностью сохранения в БД найденных контекстов.

**Принцип четвёртый — функциональность.** Разработанная БД обладает большим потенциалом, который частично реализован в рамках данного исследования. Наметим некоторые перспективы дальнейшей работы в этом направлении. Одной из перспектив является возможность реализации экспериментального подхода для получения данных о «бытовом» (в отличие от «юридического») восприятии отобранных ЛСЕ в языковом сознании обыденного носителя русского языка. Формат БД позволит как организовать подбор материала для проведения эксперимента (сформировать списки ЛСЕ для анкеты-опросника), так и внести результаты опроса в БД, добавив соответствующее поле.

Отбор ЛСЕ проводится методом сплошной выборки. Работу с материалом можно описать в два этапа: этап отбора материала и этап заполнения таблицы. При отборе материала ведущими принципами являются семантический, стилистический и тематический. При заполнении таблицы из словаря-источника извлекаются следующие данные: ЛСЕ, стилистические пометы к ним (при наличии), толкование, также делаются отметки о случаях переносного значения, приводятся дополнительные пометы «устар.» и «собир.» и комментарии. Также для каждого слова обозначается тематическая группа и указывается номер страницы из словаря-источника.

Всего изначально из словника РСС было отобрано около 2 000 единиц, из которых в конечную выборку вошло 1 829 ЛСЕ. Из этих 1 829 ЛСЕ 217 являются лексико-семантическими вариантами многозначных слов и, соответственно, повторяются в разных тематических группах, то есть «уникальными» (неповторяющимися) словами являются 1 612. В общей сложности в БД вошло из РСС:

- ЛСЕ с отрицательным денотативным значением — 1 109 (60,5 %);
- ЛСЕ с отрицательно-оценочным коннотативным значением — 720 (39,5 %): из них с пометой «ирон.» — 86, «неодобр.» — 142,

- «пренебр.» — 163, «унич.» — 13, «презр.» — 160, «груб.» — 35, «бран.» — 197, «груб. бран.» — 18, «презр. бран.» — 2;
- ЛСЕ с прямым значением — 1 592 (87 %), с переносным — 237 (13 %);
  - ЛСЕ, обозначающих лицо — 1 713, обозначающих группу лиц — 123;
  - 1 762 имени существительных, 46 идиом, приравненных к слову, а также 21 единица с неопределённым статусом, но выраженной отрицательно-оценочной семантикой номинации лица или группы лиц (типа *'Звездная болезнь* — о том, кто, добившись популярности, известности, зазнается, ведет себя неподобающим образом', *'Ни сват ни брат* (неодобр.) — о совершенно чужом, постороннем человеке' и под.).

Охарактеризуем структуру БД: её формируют несколько полей, содержащих информацию как лингвистического, так и вспомогательного характера. В поле «**Код**» указан уникальный для каждой записи код, являющийся порядковым номером ЛСЕ. Наполнение поля «**ЛСЕ**» — отрицательно-оценочные ЛСЕ, выбранные из РСС. Поле «**Отсылка**» содержит указания на наличие семантической связи с другой ЛСЕ. В поле «**РСС**» приводятся стилистические пометы из этого словаря в соответствии с принятой в нём системой помет. «**Толкование**»: толкование приводится из РСС. При многозначных словах указывается только то значение, которое выражает семантику отрицательной оценки. «**Тип значения**»: в данном поле делается отметка «перен.» в случаях непрямого значения ЛСЕ. Столбец «**Доп. информация**» заполняется информацией двух типов — хронологической пометой «устар.» и семантической пометой «собрир.». «**Комментарий**» содержит извлечённую из РСС информацию различного характера, например, о сфере употребления, об этимологии ЛСЕ. В поле «**ТГ**» указывается тематическая группа, к которой относится ЛСЕ. В поле «**Стр.**» записаны сведения о постраничном расположении ЛСЕ в словнике РСС.

Описанная в данной статье БД сформирована для автоматического поиска оценочной лексики и входит в состав разрабатываемой экспертной системы, предназначенной для лингвистической экспертизы конфликтных текстов. Данное исследование проводится 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.

## ЛИТЕРАТУРА

- Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. М.: Эдиториал УРСС, 2001. 360 с.
- Голев Н. Д., Матвеева О. Н.* Юрислингвистическая экспертиза: на стыке языка и права // Сибирский филологический журнал. № 1. Новосибирск: НГУ, 2003. С. 146–157.
- Мишанкина Н. А.* Базы данных в лингвистических исследованиях // Вопросы лексикографии. 2013. № 1 (3). С. 25–33.
- Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / Российская академия наук. Ин-т рус. яз. им. В. В. Виноградова; Под общей ред. Н. Ю. Шведовой. — Том 1: Слова указующие (местоимения). Слова именующие: имена существительные (Всё живое. Земля. Космос). М.: Азбуковник, 1998. 800 с.