

А. Б. Хертек, Б. Ч. Ооржак, А. Я. Салчак, В. С. Ондар, С. М. Далаа

ОПЫТ РАЗРАБОТКИ ЭЛЕКТРОННОЙ БАЗЫ ДАННЫХ ЛЕКСЕМ ТУВИНСКОГО ЯЗЫКА¹

Данная статья содержит информацию о первом опыте разработки электронной базы данных лексем тувинского языка в рамках проекта «Создание базы данных лексического фонда тувинского языка» (РГНФ/РФФИ № 16-04-12020, 2016-2017 гг.). Созданные базы содержат основной фонд полнзначных лексем (имена существительные, имена прилагательные, наречия, глаголы, также местоимения), распределенных по семантическим классам, подклассам, группам, подгруппам и микрогруппам. Систематизированная база данных лексического фонда будет использоваться при дальнейшей работе по семантической разметке текстов Электронного корпуса тувинского языка, составлении разных типов словарей по тувинскому языку.

Ключевые слова: *тувинский язык, база данных, лексический фонд, лексико-семантические категории, лексико-семантические классы, лексико-семантические подклассы, лексическая сочетаемость.*

Проект «Создание базы данных лексического фонда тувинского языка» (РГНФ/РФФИ № 16-04-12020, 2016-2017 гг.) является продолжением работы по разработке Электронного корпуса текстов тувинского языка (ЭКТТЯ). Электронная база данных лексического фонда тувинского языка представляет собой справочно-поисковую систему, при помощи которой будет автоматизирован поиск необходимой семантической информации из текстов ЭКТТЯ и будет служить основой для создания его семантической разметки. База данных позволит эффективно и быстро получать информацию о семантической и стилистической принадлежности лексемы, определять состав лексических единиц, формирующих определенную семантическую область.

Другой важной и значимой задачей в работе над созданием баз данных лексического фонда тувинского языка является выявление лексической сочетаемости лексем. В автоматизированную систему будут включены семантически допустимые сочетания лексем (двух существительных, прилагательных и существительных, местоимений и существительных, наречий и глаголов, существительных и глаголов). Автоматизированной системой будут исключены семантически недопустимые сочетания. Создание разных пользовательских запросов с учетом семантики позволит уточнить, выявить правила сочетаемости тех или лексических единиц.

Для создания баз данных участниками проекта проделана большая словарная работа. Из существующих словарей тувинского языка выделены полнзначные лексем (имена существительные, имена прилагательные, наречия, глаголы, также местоимения), которые составляют основной словарный фонд тувинского языка, отражающий специфику традиционных представлений носителей. Так, в материал не включаются заимствования из русского языка на тувинский, а также заимствования из европейских языков, вошедшие через русский язык.

Основой электронной базы данных послужили существующие словари по тувинскому языку: «Тувинско-русский словарь» (20 000 слов) (1955), «Тувинско-русский словарь» (около 22 000 слов) (1968, переиздание 2014), «Толковый словарь тувинского языка» (2003, 2011), «Этимологический словарь тувинского языка» (Татаринцев, 2000; 2002; 2004; 2008). Списки лексем были из лингвистических словарей были дополнены данными специальных работ по соответствующей тематике: «Топонимический словарь» (Ондар, 2004), «Шестязычный словарь названий птиц Тувы и Западной Монголии» (1999), «Определитель растений Республики Тывы» (Красноборов, 2007), «Сравнительно-историческая грамматика тюркских языков. Лексика» (2001), «Природное окружение и материальная культура пратюркских народов» (2008),

¹ Работа выполнена в соответствии с планом научно-исследовательских работ Государственного задания Министерства образования и науки Российской Федерации № 1.4539.2017/8.9 (проект № 34.3876.2017/ПЧ).

«Лошадь в традиционной практике тувинцев-кочевников» (Даржа, 2003); «Традиционные мужские занятия тувинцев» (Даржа, 2013).

На основе списков лексем созданы базы данных по именам существительным, именам прилагательным, местоимениям, наречиям и глаголам. Выбранные лексем, входящие в базовые семантические блоки, были распределены по семантическим классам, подклассам, группам, подгруппам и микрогруппам. Затем были проанализированы и приведены в соответствие переводы (толкования лексем, отражающих национально-культурное своеобразие культуры, уклада жизни и быта тувинского народа).

Разработанная электронная база данных основывается на распределении всех однозначных лексем тувинского языка на семантические разряды (классы) слов. Условно выделены пять базовых семантических классов: «Человек», «Животное», «Предмет», «Природные объекты и явления», «Абстрактные понятия», пронизывающие всю систему лексики. Далее они подразделяются на более дробные семантические подклассы (все категории, подклассы и дескрипторы в базе обозначаются тэгами на тувинском, русском и английском языках).

База данных ИМЕН СУЩЕСТВИТЕЛЬНЫХ (всего 5368 ед.) (здесь и далее в скобках даются количество лексем, относящиеся к данным группам) включает в себя следующие подклассы:

I. Предметные имена существительные включают следующие семантические группы:

1. В классе «Человек» выделяются подгруппы: имена родства (27), занятия и профессия (66), этнонимы (46). Всего лексем - 139.

2) «Животное»: дикие животные (хищники (50), травоядные (28), грызуны (30), насекомоядные (5), пресмыкающиеся (7), земноводные и моллюски (3)); домашние животные (половозрастные названия): лошади (19), корова (8), овца (9), коза (7), як (4), олень (12), верблюд (5), собака (3), кошка (4), свинья (1), осел (1), птицы (дикие птицы (110), домашние птицы (3)), рыбы (20), насекомые (29); названия частей тела животного (22) строение головы (39), названия внутренних органов животного (25), названия костей скелета (23), биол. жидкости (6), мясо, сало (14). Всего лексем – 487.

3) «Предмет»: одежда (зимняя одежда (14), демисезонная одежда (6), летняя одежда (7), детская одежда (1), одежда борца (1), головной убор (35), обувь (7); детали одежды (10), детали головного убора (4), детали обуви (2)); посуда и утварь (100), здания и сооружения (жилище (5) юрта, чум (23), дом (28)), мебель (5), предметы домашнего обихода (10), нежилые постройки (23), мужские рабочие инструменты и принадлежности (48) женские украшения и принадлежности (13) фольклорные музыкальные инструменты (10), бытовые фоноинструменты (7), обрядовые фоноинструменты шаманского камлания (8), обрядовые фоноинструменты буддийской храмовой службы (13)); снаряжения коня (33), инструменты (инструменты для охоты (19), инструменты для ловли рыбы (10), рабочие инструменты (19)); учебные принадлежности (14); продукты питания (молочные продукты (14), мясные продукты (31), растительная пища (68). Всего лексем - 578.

4) «Природные объекты и явления»: растения (травы (87), деревья (9), кустарники (18), цветы (13), ягоды (17), зерновые (14)); названия погодных явлений (66), небо и небесные тела (27), объекты ландшафта (суша (горы и возвышенности (59), лес и тайга (16), поляны (5), впадины (15), особенности поверхности земли (14), почва (15), местности (10)), водные объекты (60), погода и явления природы (54), сезоны, времена года (4), время суток (8), небо и небесные объекты (27). Всего лексем - 528.

II. Имена собственные состоят из семантических классов:

1) «Человек»: антропонимы: личные женские имена (105), мужские имена (108). Всего лексем – 213.

2) «Природные объекты»: топонимы: гидронимы (названия рек (1071), озер (158), минеральных источников (63), омутов (5); оронимы: названия гор (586), скал (6), хребтов (109),

степей (4) и таежной местности (10); ойконимы: названия районов (15), городов (5), населенных пунктов (135) и местечек (987). Всего лексем – 3154.

3) «Животное»: (клички собак (22), лошадей (78), коров (29) и овец, коз (22). Всего лексем – 151.

III. Непредметные имена существительные входят в класс «Абстрактные понятия»: эмоции (54), чувственные восприятия (20), универсальные представления (44). Всего лексем – 118.

База данных ИМЕН ПРИЛАГАТЕЛЬНЫХ включает в себя 1277 имен прилагательных.

Прилагательные были распределены на группу качественных и относительных прилагательных по грамматическому признаку.

В классе «Человек» среди качественных прилагательных выделены следующие семантические группы прилагательных: «Физические качества» (165) с подгруппами «Тело, телосложение» (29), «Рост» (7), «Внешность» (25), «Баш» (4), «Волосы» (10), «Лицо» (14), «Глаза» (15), «Брови» (3), «Уши» (3), «Нос» (5), «Щеки» (4), «Губы» (4), «Зубы» (3), «Лоб» (2), «Руки» (7), «Пальцы» (5), «Спина» (5), «Ноги» (7), «Кожа» (6), «Живот» (7); «Физиологические качества» (25); «Умственные качества» с подгруппами «Положительные качества» (13), «Отрицательные качества» (7); «Психические качества и характер» с подгруппами «Положительные качества» (41), «Отрицательные качества» (115); «Душевные качества» с подгруппами «Положительные качества» (30), «Отрицательные качества» (18); «Нравственные качества» с подгруппами «Положительные качества» (11), «Отрицательные качества» (14); «Социальный признак» с подгруппами «Возрастной признак» (12), «Имущественный признак» (15), «Признание / не признание в обществе, семье, коллективе» (26), «Способности, таланты» (19).

В классе «Человек» (77) среди относительных прилагательных выделены следующие семантические группы прилагательных: «Возраст» (4); «Физиологическое состояние» (15); «Семейное положение» (4); «Эмоциональное состояние» (10); Имущество (15); «Обладающий/Не обладающий каким-л. качеством» (29).

В классе «Природные объекты и явления» выделены следующие семантические группы прилагательных: «Погода» (23); «Ландшафт. Суша» (20); «Ландшафт. Водные объекты» (18).

В классе «Животное» (56) выделены следующие семантические группы прилагательных: «Физические качества» (30); «Физиологические качества» (6); «Поведение» (10); «Мать» (52).

В классе «Предмет» (238) среди качественных прилагательных выделены следующие семантические группы прилагательных: «Размер» (13); «Физическое качество» (29); «Количество» (11); «Форма» (30); «Цвет» (19); «Вес» (3); «Вкус» (11); «Запах» (10); «Температура» (8); «Звук» (15); «Положительные качества» (37); «Отрицательные качества» (52).

В классе «Предмет» (292) среди относительных прилагательных выделены следующие семантические группы прилагательных: «Материал» (9); «Время» (22); «Расстояние» (4); «Отношение» (10); «Количество» (6); Физические свойства (29); «Эмоции» (4); «Имеющий к.-л. качество» (208).

База данных МЕСТОИМЕНИЙ включает: личные местоимения (6), указательные (42), вопросительные (53), определительные (24), неопределенные (16). Всего лексем - 141.

Составленная база данных СОБСТВЕННО НАРЕЧИЙ состоит из 260 лексем. Наречия были распределены в лексико-грамматические группы: наречия меры и степени (23); наречия образа действия (90); временные наречия (103); пространственные наречия (40); наречия цели (4).

Базы данных ГЛАГОЛОВ² (всего 2237 ед.), входящих в следующие тематические блоки и внутренние семантические группы:

² Семантическая классификация глаголов тувинского языка проводится в соответствии с предложенной классификацией в «Большом толковом словаре русских глаголов: Идеографическое описание. Английские эквиваленты. Синонимы. Антонимы» (Бабенко, 2007).

I. **ДЕЙСТВИЕ И ДЕЯТЕЛЬНОСТЬ**: 1) движение; 2) перемещение объекта; 3) помещение; 4) физическое воздействие на объект; 5) созидательная деятельность; 6) интеллектуальная деятельность; 7) поведения; 8) речевая деятельность; 9) физиологическое действие; 10) звучание.

II. **БЫТИЕ И КАЧЕСТВЕННОЕ СОСТОЯНИЕ**: 1) бытие; 2) качественное состояние.

III. **ОТНОШЕНИЕ**: 1) взаимоотношение; 2) владение; 3) социальные отношения.

Самым обширным тематическим блоком глаголов является **ДЕЙСТВИЕ И ДЕЯТЕЛЬНОСТЬ** (всего 1364 ед.), в который включаются:

Группа глаголов **ДВИЖЕНИЯ** (всего 246 ед.), распределенная на две подгруппы: 1) глаголы направленного движения; 2) глаголы ненаправленного движения. Подгруппа глаголов *направленного движения* представлена пятнадцатью микрогруппами: глаголы субъектной ориентации движения (6); глаголы с семой 'достижения конечного пункта или приближения к конечному пункту' (12); глаголы пространственной ориентации движения (4); глаголы, образованные от указательных местоимений (2); глаголы с семой 'удаления от исходного пункта' (11); глаголы с семой 'вхождения' (4);

Глаголы с семой 'покидания объекта' (2); глаголы с семой 'направления движения' (5); глаголы со значением 'двигаться с указанием цели движения' (15); глаголы, содержащие сему 'трасса передвижения' (20); глаголы со значением 'идти за чем-либо' (29); глагол со значением 'ехать на чем-либо' (1), глаголы со значением движения 'вверх' (12); глаголы со значением движения 'вниз' (28). Подгруппа глаголов *ненаправленного движения* представлена десятью микрогруппами: глаголы с семой движения 'по твердой поверхности' (25); глаголы с семой движения в 'жидкой среде' (4); глаголы с семой 'движения по воздуху' (6); глаголы интенсивного темпа движения (12); глаголы медленного темпа движения (3); глаголы с семантикой 'преодоления препятствия' (4); глаголы с семой образного движения (21); глаголы со значением 'расходиться в разные стороны' (3); глаголы с семантикой 'следования' (6); глаголы с семантикой 'вокруг чего-либо' (11).

Группа глаголов **ПЕРЕМЕЩЕНИЯ** (всего 168 ед.) подразделяется на две подгруппы: глаголы направленного перемещения (138) со следующими микрогруппами: глаголы однонаправленного перемещения,

ориентированного относительно исходного и конечного пунктов (30), глаголы однонаправленного перемещения, ориентированного относительно исходного пункта (51), глаголы однонаправленного перемещения, ориентированного относительно конечного пункта (48), глаголы однонаправленного перемещения, ориентированного относительно промежуточного пункта (9); глаголы ненаправленного перемещения - глаголы кругового и вращательного перемещения (30).

Группа глаголов **ПОМЕЩЕНИЯ ОБЪЕКТА** (всего 194 ед.), имеют в своем составе подгруппы глаголов: помещения объекта в определенном месте каким-либо образом (38), помещения объекта в результате физического действия (76), включения объекта в состав чего-либо (5), собственно покрытия объекта (20).

Группа глаголов **ФИЗИЧЕСКОГО ВОЗДЕЙСТВИЯ НА ОБЪЕКТ** (всего 190 ед.) делится на 8 подгрупп: нанесения удара (30), давления (8), повреждения неодушевленного объекта (19), повреждения тела живого существа (25), прикосновения (28), соединения и присоединения (34), разделения и отделения (22), очищения и удаления объекта (24).

Группа глаголов **СОЗИДАТЕЛЬНАЯ ДЕЯТЕЛЬНОСТЬ** (всего 110 ед.) состоит из глаголов создания объекта в результате трудовой деятельности, глаголов приготовления чего-либо, глаголов собирания, глаголов графической передачи информации. Глаголы создания объекта в результате трудовой деятельности распределены на две подгруппы: 1) глаголов создания объекта в результате физического труда (42); 2) глаголов создания объекта в результате интел-

лектуального труда (20). В группе глаголов приготовления чего-либо выделена подгруппа глаголов глаголы приготовления пищи (19). Здесь же выделены глаголы собирания (19), глаголы графической передачи информации (10).

Группа глаголов **ИНТЕЛЛЕКТУАЛЬНОЙ ДЕЯТЕЛЬНОСТИ** (всего 44 ед.) распределены на следующие подгруппы: глаголы восприятия (6); понимания (16); познания (4); мышления (11); сравнения и сопоставления (3); выбора (3).

Группа глаголов **ПОВЕДЕНИЯ** (всего 205 ед.) распределена на подгруппы: неестественного поведения (45); несерьезного поведения (21); неискреннего поведения (30); нечестного поведения (27); безответственного поведения (9); бесцеремонного поведения (18); самовольного поведения (15); безнравственного поведения (10); неуравновешенного поведения (30).

Группа глаголов **РЕЧЕВОЙ ДЕЯТЕЛЬНОСТИ** (всего 113 ед.) по семантическому признаку распределены в следующие подгруппы: глаголы речевого сообщения (44), глаголы речевого общения (25), глаголы речевого воздействия (23), глаголы характеризованной речевой деятельности (21).

Группа глаголов **ФИЗИОЛОГИЧЕСКОГО ДЕЙСТВИЯ** составляют 26 глаголов.

Группа глаголов **ИЗМЕНЕНИЯ ПОЛОЖЕНИЯ В ПРОСТРАНСТВЕ** включает 14 лексем.

Группа **ГЛАГОЛОВ ЗВУЧАНИЯ** (всего 54 ед.).

Тематический блок **БЫТИЕ И КАЧЕСТВЕННОЕ СОСТОЯНИЕ**

(всего 659 ед.):

Группа глаголов **БЫТИЕ** (всего 214 ед.) распределена в следующие подгруппы: глаголы начальной фазы бытия, существования (35), глаголы

существования (72), глаголы прекращения бытия (98), глаголы образа жизни (9).

Группа глаголов **КАЧЕСТВЕННОЕ СОСТОЯНИЕ** (всего 445 ед.) распределена в следующие подгруппы: глаголы становления качества со следующими микрогруппами: глаголы становления физического качества (44), глаголы становления и проявления цветового признака (32), глаголы становления внешних признаков человека (17), глаголы становления внутренних признаков человека (17), глаголы становления внешних признаков животного (2), глаголы становления внутренних признаков животного (3), глаголы становления социальных качеств (2); глаголы проявления признака со следующими микрогруппами: глаголы проявления качества (37), глаголы изменения качественного признака (21), глаголы изменения количественного признака (24); глаголы эмоционального состояния: глаголы пребывания субъекта в определенном эмоциональном состоянии (95), глаголы становления эмоционального состояния (51), глаголы приведения в эмоциональное состояние (100).

Тематический блок **ОТНОШЕНИЕ** (всего 214 ед.) включает три группы глаголов: взаимоотношение, владение, социальные отношения.

Группа глаголов **ВЗАИМООТНОШЕНИЕ** (всего 31 ед.) распределена в две подгруппы: глаголы взаимосвязи (21), глаголы замены (10).

Группа глаголов **ВЛАДЕНИЕ** (всего 97 ед.) распределена на 2 подгруппы: глаголы поиска объекта (18); глаголы приобретения с микрогруппами: глаголы получения объекта в определенном количестве (5), глаголы получения в свое распоряжение (15), глаголы отчуждения (11), глаголы компенсированного приобретения (6), глаголы обладания (2), глаголы сохранения (5), глаголы утраты объекта (7), глаголы лишения (9), глаголы передачи объекта (19).

Группа глаголов **СОЦИАЛЬНЫЕ ОТНОШЕНИЯ** (всего 86 ед.) распределена в следующие подгруппы: глаголы победы и поражения (12), глаголы принуждения (13), глаголы собственно влияния (20), глаголы убеждения (5), глаголы подчинения (3), глаголы защиты (6), глаголы помощи (5), глаголы обеспечения (8), глаголы разрешения и запрещения (4), глаголы управления (10).

Составлены **БАЗЫ ДАННЫХ ПО СОЧЕТАЕМОСТИ ГЛАГОЛОВ** с именами существительными и наречиями.

Сочетаемость глаголов с именами существительными

Базы данных по сочетаемости глаголов ДВИЖЕНИЯ, ПЕРЕМЕЩЕНИЯ ЗВУЧАНИЯ, БЫТИЯ, КАЧЕСТВЕННОЕ СОСТОЯНИЕ с именами существительными показывают, что все они употребляются имена существительные, которые относятся к классам «Человек», «Животное», «Предмет», «Природные явления», так как они описывают действия, характерные как для живого существа, так и неживого.

База данных по сочетаемости глаголов, входящих в группу СОЗИДАТЕЛЬНАЯ ДЕЯТЕЛЬНОСТЬ, ИНТЕЛЛЕКТУАЛЬНАЯ ДЕЯТЕЛЬНОСТЬ, ГЛАГОЛОВ ПОВЕДЕНИЯ, РЕЧЕВОЙ ДЕЯТЕЛЬНОСТИ, ВЗАИМООТНОШЕНИЕ, СОЦИАЛЬНЫЕ ОТНОШЕНИЯ и ПОМЕЩЕНИЕ ОБЪЕКТА показывают, что с глаголами данных групп употребляются имена существительные, которые относятся к классу «Человек», так как глаголы данной группы описывают действия, характерные только для человека. У глагола *чанна* = 'вести себя' (группа поведения) возможна сочетаемость с существительным *кудай* 'небо'. Субъектом поведения при данном глаголе могут выступать домашние животные

С существительными классов «Человек» и «Животное» сочетаются глаголы групп ФИЗИЧЕСКОГО ВОЗДЕЙСТВИЯ НА ОБЪЕКТ, ФИЗИОЛОГИЧЕСКОГО ДЕЙСТВИЯ, ВЛАДЕНИЯ.

Сочетаемость глаголов с наречиями

Наречия, обозначающие ОБРАЗ ДЕЙСТВИЯ сочетаются с глаголами групп: движения, физического воздействия на объект, физиологического действия, изменения положения в пространстве, созидательной деятельности, интеллектуальной деятельности, глаголов поведения, речевой деятельности, бытия, социальные отношения, взаимоотношения, владения.

Наречия, характеризующие МЕРУ И СТЕПЕНЬ ПРОЯВЛЕНИЯ ПРИЗНАКА ДЕЙСТВИЯ употребляются с группами глаголов звучания, физиологического действия, речевой деятельности, бытия, качественного состояния, владения.

ПРОСТРАНСТВЕННЫЕ наречия сочетаются с глаголами движения, перемещения, помещения объекта, изменения положения в пространстве, бытия.

ВРЕМЕННЫЕ наречия употребляются с глаголами движения, речевой деятельности, бытия.

Созданные базы содержат основной фонд исконных 7046 именных и 2237 глагольных лексем тувинского языка, которые будут использованы при дальнейшей работе по семантической разметке текстов Электронного корпуса тувинского языка. На их основе будет создаваться программа поиска допустимых/не допустимых сочетаний лексем. После разработки данной программы и ее апробации составленные базы данных будут по необходимости скорректированы и дополнены.

Электронная база данных лексического фонда тувинского языка открывает новые возможности в исследовательской работе в области лексической семантики, а также в лексикографии. Она ускоряет поиск необходимой информации, что экономит время исследователя и расширяет географию пользователей; упрощает работу при создании различных словарей. Базы данных лексического фонда будут использоваться для составления разных типов словарей по тувинскому языку: тематических, объяснительных, словарей сочетаемости, словарей синонимов и антонимов, частотных словарей и т. д.

Создание электронных баз производится с помощью системы управления баз данных Access 2010. Тексты на тувинском языке будут обрабатываться с помощью системы объектно-ориентированного программирования C++. Эти системы поддерживают кодировку Unicode, в которой оцифрованы все тексты на тувинском языке. Будут создаваться компьютерные программы как для компьютеров с операционной системой Windows, так и для мобильных устройств с операционной системой Android. В настоящее время поиск производится в программе Excel. Информация о создании баз данных по проекту размещена в сети Интернет на странице Электронного корпуса текстов тувинского языка <http://tuvancorpus.ru/?q=content/bazy-dannyh>

Список литературы:

- Даржа В. К.** Лошадь в традиционной практике тувинцев-кочевников.– Кызыл: ТИКОПР СО РАН. 2003.
Даржа В. К. Традиционные мужские занятия тувинцев. – Кызыл: Тываполиграф. 2013.
Красноборов И М., Ломоносова М. Н., Шаула Д. Н. и др. Определитель растений Республики Тывы. – Новосибирск: Изд-во СО РАН, 2007.
Сравнительно-историческая грамматика тюркских языков. Лексика. – М.: Наука, 2001.
Природное окружение и материальная культура пратюркских народов. – М.: Вост. Литература, 2008.

Словари:

- Бабенко Л. Г.** Большой толковый словарь русских глаголов: Идеографическое описание. Английские эквиваленты. Синонимы. Антонимы – М: АСТ-Пресс, 2007.
Ондар Б. К. Топонимический словарь. – Абакан: Изд-во ХГУ. 2004.
Татаринцев Б. И. Этимологический словарь тувинского языка. – Новосибирск: Наука, 2000. – Т. I. – 344 с. – 2002. – Т. II. – 388 с.; 2004. – Т. III – 440 с.; 2008. – Т. IV.
Тувинско-русский словарь. М.: Гос. Изд-во иностранных и национальных словарей. 1955.
Тувинско-русский словарь. М.: Сов. энциклопедия. 1968. – Переиздание. Кызыл: Тув. книжное изд-во. 2014.
Толковый словарь тувинского языка. – Новосибирск: Наука, 2003. – Т. I. – 596 с.; 2011. – Т. II.
Шестязычный словарь названий птиц Тувы и Западной Монголии (сост. В. И. Забелин). – Новосибирск: Изд-во СО РАН. 1999.

Хертек Аржаана Борисовна, кандидат филологических наук, старший научный сотрудник.

Научно-образовательный центр «Тюркология»

Тувинского государственного университета.

Ул. Ленина, 36, Кызыл, Республика Тыва, Россия, 667000.

E-mail: khertek.ab@yandex.ru

Ооржак Байлак Чаш-ооловна, доктор филологических наук, старший научный сотрудник.

Научно-образовательный центр «Тюркология»

Тувинского государственного университета.

Ул. Ленина, 36, Кызыл, Республика Тыва, Россия, 667000.

E-mail: oorzhak.baylak@mail.ru

Салчак Аэлита Яковлевна, кандидат филологических наук, старший научный сотрудник.

Научно-образовательный центр «Тюркология»

Тувинского государственного университета.

Ул. Ленина, 36, Кызыл, Республика Тыва, Россия, 667000.

E-mail: aelita_74@mail.ru

Ондар Валентина Сувановна

Научно-образовательный центр «Тюркология»

Тувинского государственного университета.

Ул. Ленина, 36, Кызыл, Республика Тыва, Россия, 667000.

E-mail: barys-hoov@mail.ru

Далаа Сергей Монгушевич, кандидат физико-математических наук, старший научный сотрудник.

Научно-образовательный центр «Тюркология»

Тувинского государственного университета.

Ул. Ленина, 36, Кызыл, Республика Тыва, Россия, 667000.

E-mail: dalaa@rambler.ru

Материал поступил в редакцию 5 ноября 2019 г.

A. B. Hertek, B. C. Oorzhak, A. Y. Salchak, V. S. Ondar, S. M. Dalaa

EXPERIENCE IN DEVELOPING AN ELECTRONIC DATABASE OF TUVAN LEXEMES

This article contains information about the first experience in the development of an electronic database of Tuvan lexemes within the framework of the project "Creation of a database of the Tuvan lexical Fund" (RGNF / RFBR No. 16-04-1220, 2016-2017). The created databases contain the main body of full lexemes (nouns, adjectives, adverbs, verbs, pronouns), distributed by semantic classes, subclasses, groups, subgroups, and microgroups. Systematized database of lexicon will be used for further work on semantic markup of electronic texts corpus of the Tuvan language, compiling various types of dictionaries for Tuvan language. Databases will be used to compile different types of dictionaries.

Creation of electronic databases is performed using the Access2010 database management system. Texts in the Tuvan language will be processed using the C ++ object-oriented programming system. These systems support Unicode encoding, in which all texts in Tuvan are digitized. Computer programs will be created both for computers with the Windows operating system, and for mobile devices with the Android operating system. Currently, the search is performed in the program Excel. Information on the creation of databases for the project is available on the Internet on the page of the Electronic Corpus of Tuvan Language Texts <http://tuvancorpus.ru/?q=content/bazy-dannyh>.

Key words: *Tuvan language, database, lexicon, lexical-semantic categories, lexical-semantic classes, lexical-semantic subclasses, lexical compatibility.*

References:

- Darja V. K.** Horse in the traditional practice of Tuvan nomads. – Kyzyl: TUVIKOPR SB RAS. 2003. (in Russian)
Darja V. K. Traditional male occupations of Tuvinians. – Kyzyl: Typeparagraph. 2013. (in Russian)
Krasnoborov and M., Lomonosova M. N., Shaulo D. N. and others. Determinant of plants of the Republic of Tyva. – Novosibirsk: SB RAS publishing House, 2007. (in Russian)
Comparative-historical grammar of Turkic languages. Vocabulary. – Moscow: Nauka, 2001. (in Russian)
Natural environment and material culture of the pre-Turkic peoples. – M.: East. Literature, 2008. (in Russian)

Dictionaries:

- Babenko L. G.** Big explanatory dictionary of Russian verbs: Ideographic description. English equivalent. Synonyms. Antonyms. – M: AST Press, 2007. (in Russian)
Ondar B. K. Toponymic dictionary. – Abakan: Publishing house of KSU. 2004. (in Russian)
Tatarintsev B. I. Etymological dictionary of the Tuvan language. – Novosibirsk: Nauka. T. I. 2000. 344 S. T. II, 2002. 388 S. T. III, 2004. 440 p.; IV, 2008. (in Russian)
Tuvan-Russian dictionary. – Moscow: State Publishing house of foreign and national dictionaries. 1955. (in Russian)
Tuvan-Russian dictionary. M.: The Scoop. encyclopedia. 1968. – Reissue. Now: TUV. in the bookstore-the root. 2014.
Explanatory dictionary of the Tuvan language. – Novosibirsk: Nauka. T. I. 2003. - 596 S. T. II, 2011. (in Russian)
Six-language dictionary of bird names of Tuva and Western Mongolia (comp. Zabelin V. I.). Novosibirsk: publishing house of SB RAS. 1999. (in Russian)

Hertek Arzhaana Borisovna, Candidate of philological Sciences, senior researcher.

Scientific-educational center "Turkology" Tuvan state University.

Ul. Lenina, 36, Kyzyl, Republic of Tyva, Russia, 667000.

E-mail: khertek.ab@yandex.ru

Oorzhak Baylak Chash-oolovna, Doctor of Philology, senior researcher.

Scientific-educational center "Turkology" Tuvan state University.

Ul. Lenina, 36, Kyzyl, Republic of Tyva, Russia, 667000.

E-mail: oorzhak.baylak@mail.ru

Salchak Aelita Yakovlevna, Candidate of philological Sciences, senior researcher.

Scientific-educational center "Turkology" Tuvan state University.

Ul. Lenina, 36, Kyzyl, Republic of Tyva, Russia, 667000.

E-mail: aelita_74@mail.ru

Ondar Valentina Suvanova, Candidate of philological Sciences, senior researcher.

Scientific-educational center "Turkology" Tuvan state University.

Ul. Lenina, 36, Kyzyl, Republic of Tyva, Russia, 667000.

E-mail: barys-hoov@mail.ru

Dallaa Sergey Mongushevich, Candidate of physical and mathematical Sciences, senior researcher.

Scientific-educational center "Turkology" Tuvan state University.

Ul. Lenina, 36, Kyzyl, Republic of Tyva, Russia, 667000.

E-mail: dalaa@rambler.ru