

СЕМАНТИЧЕСКАЯ РАЗМЕТКА ИМЕНИ ДЛЯ ЭЛЕКТРОННОГО КОРПУСА ТЕКСТОВ ТУВИНСКОГО ЯЗЫКА

*Байлак Ч. Ооржак,
Аржаана Б. Хертек,
Мария А. Кужугет,
Валентина С. Ондар*
Тувинский государственный университет

SEMANTIC MARKUP OF NOUNS AND ADJECTIVES FOR THE ELECTRONIC CORPUS OF TEXTS IN TUVAN LANGUAGE

*Bajlak Ch. Oorzhak
Arzhaana B. Khertek
Marija A. Kuzhuget
Valentina S. Ondar*
Tuvan State University



Работа выполнена при поддержке РГНФ (проект «Создание базы данных лексического фонда тувинского языка», грант №16-04-12020).

Ооржак Байлак Чаш-ооловна — кандидат филологических наук, старший научный сотрудник Научно-образовательного центра «Тюркология» Тувинского государственного университета. Адрес: 667000, Россия, г. Кызыл, ул. Ленина, д. 32. Тел.: +7 (394-22) 3-03-78. Эл. адрес: oorzhak.baylak@mail.ru

Хертек Аржаана Борисовна — кандидат филологических наук, старший научный сотрудник Научно-образовательного центра «Тюркология» Тувинского государственного университета. Адрес: 667000, Россия, г. Кызыл, ул. Ленина, д. 32. Тел.: +7 (394-22) 3-03-78. Эл. адрес: khertek.ab@yandex.ru

Кужугет Мария Амын-ооловна — заведующая литературным музеем Тувинского государственного университета. Адрес: 667000, Россия, г. Кызыл, ул. Ленина, д. 32. Тел.: +7 (394-22) 3-10-62. Эл. адрес: kuzhuget.m55@mail.ru

Ондар Валентина Сувановна — кандидат филологических наук, доцент кафедры русского языка и литературы Тувинского госуниверситета. Адрес: 667000, Россия, г. Кызыл, ул. Ленина, д. 32. Тел.: +7 (394-22) 5-22-50. Эл. адрес: barys-hoov@mail.ru

Oorzhak Bajlak Chash-oolovna, Candidate of Philology, Senior Research Fellow, Research and Education Center of Turkic Studies, Tuvan State University. Postal address: 32 Lenin St., 667000 Kyzyl, Republic of Tuva, Russian Federation. Tel: +7 (394-22) 3-03-78. E-mail: oorzhak.baylak@mail.ru

Khertek Arzhaana Borisovna, Candidate of Philology, Senior Research Fellow, Research and Education Center of Turkic Studies, Tuvan State University. Postal address: 32 Lenin St., 667000 Kyzyl, Republic of Tuva, Russian Federation. Tel: +7 (394-22) 3-03-78. E-mail: khertek.ab@yandex.ru

Kuzhuget Marija Amyn-oolovna, Head of the Literature Museum, Tuvan State University. Postal address: 32 Lenin St., 667000 Kyzyl, Republic of Tuva, Russian Federation. Tel: +7 (394-22) 3-10-62. E-mail: kuzhuget.m55@mail.ru

Ondar Valentina Suvanovna, Candidate of Philology, Associate Professor, Department of Russian language and literature, Tuvan State University. Postal address: 32 Lenin St., 667000 Kyzyl, Republic of Tuva, Russian Federation. Tel: +7 (394-22) 5-22-50. E-mail: barys-hoov@mail.ru



В статье представлен ход работ над составлением семантической разметки Электронного корпуса текстов тувинского языка (ЭКТТЯ). Этот этап является продолжением работы коллектива авторов Тувинского государственного университета (Научно-образовательного центра «Тюркология» в сотрудничестве с кафедрой информатики) по включению текстов на тувинском языке в электронную базу и разработке разметки корпуса.

Семантическая разметка лексического фонда тувинского языка будет представлять собой справочно-поисковую систему, позволяющую находить в текстах ЭКТТЯ необходимые для целей пользователя фрагменты текстов с искомым значением.

Первым шагом данного этапа работы является создание баз данных лексем тувинского языка, основанной на распределении всех полнозначных лексем тувинского языка на основные семантические классы: человек, животное, предмет, природные объекты и явления, абстрактные понятия. Все имена существительные тувинского языка, называющие объекты, а также все прилагательные (качественные и относительные), указывающие на признаки, распределяются по выделенным лексико-семантическим классам. Лексико-семантическим классам, подклассам и дескрипторам присваиваются тэги на тувинском, русском и английском языках, при помощи которых будет производиться автоматизированный поиск.

Создаваемые базы данных полнозначных лексем тувинского языка будут служить для выявления также лексической сочетаемости лексем. Предполагается, что в автоматизированной системе будет содержаться информация о семантической сочетаемости имен прилагательных и имен существительных; наречий и глаголов; имен существительных и глаголов и о невозможности семантически недопустимых сочетаний.

Ключевые слова: тувинский язык; электронная база данных; автоматизированная система поиска; лексический фонд; лексико-семантические классы и подклассы; дескрипторы; тэги; имя существительное; имя прилагательное; лексическая сочетаемость

The article examines the progress of semantic markup of the Electronic corpus of texts in Tuvan language (ECTTL), which is another stage of adding Tuvan texts to the database and marking up the corpus. ECTTL is a collaborative project by researchers from Tuvan State University (Research and Education Center of Turkic Studies and Department of Information Technologies).

Semantic markup of Tuvan lexis will come as a search engine and reference system which will help users find text snippets containing words with desired meanings in ECTTL.

The first stage of this process is setting up databases of basic lexemes of Tuvan language. All meaningful lexemes were classified into the following semantic groups: humans, animals, objects, natural objects and phenomena, and abstract concepts. All Tuvan object nouns, as well as both descriptive and relative adjectives, were assigned to one of these lexico-semantic classes. Each class, sub-class and descriptor is tagged in Tuvan, Russian and English; these tags, in turn, will help automatize searching.

The databases of meaningful lexemes of Tuvan language will also outline their lexical combinations. The automatized system will contain information on semantic combinations of adjectives with nouns, adverbs with verbs, nouns with verbs, as well as on the combinations which are semantically incompatible.

Keywords: Tuvan language; electronic database; automated search system; lexis; lexico-semantic classes and subclasses; descriptor; tag; noun; adjective; lexical compatibility

The work was performed with financial support from the Russian Foundation for the Humanities (Grant No. 16-04-12020, project title "Setting up a database of Tuvan lexis").

Введение

С 2011 г. на базе Научно-образовательного центра (НОЦ) «Тюркология» Тувинского государственного университета (ТувГУ) при поддержке кафедры информатики ТувГУ начата большая работа по применению информационных технологий к исследованиям по тувинскому языку и литературе. Проект «Электронный

корпус текстов тувинского языка» был поддержан Российским гуманитарным научным фондом (грант № 110412073).

В рамках данного проекта были оцифрованы тексты тувинской художественной литературы современного и советского периода, фольклорные тексты, тексты официально-деловых документов, словари; составлены базы данных грамматических форм имени и глагола; произведена морфологическая разметка, позволяющая автоматизировать поиск морфем и грамматических форм в заданном тексте для проведения исследовательских работ; а также был создан сайт электронного корпуса тувинского языка (www.tuvasocorpus.ru). В настоящее время база данных текстов на тувинском языке продолжает пополняться и содержится в базе данных около 300 наименований книг прозы и поэзии на тувинском языке, литературно-критические статьи, а также Морфемно-орфографический словарь тувинского языка (Бавуу-Сюрюн, Далаа, Электр. ресурс) и текст Конституции Республики Тыва на тувинском языке.

В связи с преподаванием тувинского языка и литературы в вузе и школе, подготовкой учебно-методических разработок и учебников по тувинскому языку и литературе и, связанной с ними необходимостью работы с большим объемом текстов, назрела необходимость разработки семантической разметки для текстов, включенных в ЭКТТЯ. Поэтому параллельно с морфологической разметкой началась работа по разработке семантической разметки корпуса. Предлагаемая статья коллектива авторов посвящена принципам построения семантической разметки ЭКТТЯ.

Создание баз данных и семантическая разметка имени для ЭКТТЯ

Семантическая разметка электронной базы данных лексического фонда будет функционировать в ЭКТТЯ как справочно-поисковая система, при помощи которой будет автоматизирован поиск необходимых фрагментов текстов с искомой семантической информацией с последующим их использованием при составлении учебников (в том числе и электронных) и других учебных материалов и контента, а также при составлении словарей. Первым шагом этой работы является создание баз данных лексем тувинского языка.

Разрабатываемая база данных основывается на распределении всех однозначных лексем тувинского языка на семантические разряды (классы) слов. Выделяются четыре базовые семантические классы: 1) человек, 2) животное, 3) предмет, 4) природные объекты и явления, 5) абстрактные понятия.

При этом, имена существительные подразделяются на подклассы: 1) предметные имена существительные; 2) имена собственные; 3) непредметные имена существительные (абстрактные понятия). Все предметные имена, в свою очередь,



подразделяются на: «человек» — имена родства, профессия, этноним; «животное» — дикие животные, домашние животные, птицы — дикие птицы, домашние птицы, рыбы, насекомые; «предмет» — вещества и материалы, бытовые принадлежности, здания и сооружения, инструменты; «природные объекты и явления» — растения, названия погодных явлений, небесные тела, объекты ландшафта; «абстрактные понятия» — эмоции, чувственные восприятия, универсальные представления. Далее выделенные подклассы подразделяются на более мелкие семантические группы.

Пример распределения предметных имен существительных тувинского языка и их помет приведен в таблице 1. Названия лексико-семантических классов, подклассов и дескрипторов обозначаются тэгами на тувинском, русском и английском языках.

Таблица 1. Имена существительные тувинского языка. Предметные имена.

Table 1. Nouns in Tuvan language. Object names.

Кижип / Человек / Human	Дөргүл-төрөл аттары / Имена родства / Names of kinship	ада, ача 'отец, папа', ава 'мама', кырган-ава 'бабушка', кырган-ача 'дедушка', угба 'сестра', акы 'брат', дуңма 'младший брат / младшая сестра', даай 'дядя', күүй 'жена дяди'	
	Профессия / Profession	эмчи 'врач', башкы 'учитель', ыраажы 'певец', чолаачы 'водитель, шофер'	
	Этноним / Ethnonym	кыдат 'китаец, китаянка', бурят 'бурят, бурятка', моол 'монгол, монголка', орус 'русский, русская'	
Дириг амытан / Животное / Animal	Дириг амытан / Животные / Animal	черлик / дикие	адыг 'медведь', диинч 'белка'
		азырал / домашние	инек 'корова', ыт 'собака'
	Куштар / Птицы / Birds	черлик / дикие	хартыга 'коршун', уккушкаш 'ремель'
		азырал / домашние	дагаа 'курица', кас 'гусь'
	Балыктар / Рыбы / Fishes	ак-балык 'елец', шортан 'щука', кадыргы 'хариус'	
	Курт аймаа / Насекомые / Insects	шартылаа 'кузнечик', ары 'пчела', ымыраа 'комар', сээк 'муха', шыйлашкын 'дождевой червь'	
Чүүл / Предмет / Thing	Бүдүмелдер / Вещества и материалы / Substances and materials	суг 'вода', чугай 'известь', торгу 'шелк', алдын 'золото', каң 'сталь', хөмүр 'уголь', кидис 'войлок', маны 'мрамор', хүлер 'бронза', хола 'медь, жёлтая медь', мөңгүн 'серебро'	
	Эт-херексел / Бытовые принадлежности, утварь / Household accessories	Аяк-сава / Посуда / Dishes	аяк 'пиала', паш 'чугунная чаша для приготовления пищи', диизе 'блюдец', бижек 'нож', хууң 'ведро'
		Өг, бажың дерии / Мебель / Furniture	аптара 'сундук', орун 'кровать', сандай 'табуретка, стул'



	Идик-Хеп / Одежда, обувь / Clothing, footwear	<i>тон</i> 'национальная верхняя одежда, пальто', <i>хөйлең</i> 'рубашка', <i>чүвүр</i> 'штаны, брюки', <i>идик</i> 'национальная обувь, сапоги', <i>чеңижок</i> 'безрукавка, жилетка'.	
	Бажың-балгат, тудуг объектилери / Здания и сооружения / Buildings and constructions	<i>Чурттаар оран-сава</i> / Жилище / Housing	<i>өг</i> 'юрта', <i>чадыр</i> 'чум, шалаш', <i>алажы өг</i> 'чум', <i>кидис өг</i> 'войлочная юрта', <i>доорбаш өг</i> 'бревенчатая юрта', <i>бажың</i> 'дом, здание'
		<i>Тудуг объектилери</i> / Нежилые постройки / Constructions	<i>көвүрүг</i> 'мост', <i>кажаа</i> 'загон для скота', <i>дес кажая</i> 'закрытый загон для скота'
	Эдилел / Принадлежности / Accessories	<i>Эр кижиниң эдилели</i> / Мужские принадлежности / Men's accessories	<i>балды</i> 'топор', <i>сырткыыш</i> 'удочка', <i>кестик</i> 'ножик', <i>ча</i> 'лук', <i>чаныы</i> 'брусок, оселок, мягкий точильный камень', <i>хол хирээзи</i> 'ножовка'
<i>Кыс кижиниң эдилели</i> / Женские принадлежности / Women's accessories		<i>баш шүүрү</i> 'гребень', <i>билзек</i> 'кольцо', <i>билектээш</i> 'браслет', <i>боошкун</i> 'девичье наконное украшение из трех нитей бус', <i>ине хавы</i> 'футляр для иглы', <i>сырга</i> 'серьги'	
Бойдус объектилери болгаш бойдуштуң болушкуннары / Природные объекты и явления / Natural objects and phenomena	Үнүш / растения / Plants	<i>өгт-сиген</i> 'трава', <i>ыяш</i> 'дерево', <i>шиви</i> 'ель', <i>чечек</i> 'цветок'	
	Агаар байдалы / Погодные явления / Weather conditions	<i>чаџс</i> 'дождь', <i>кызаңнаашкын</i> 'гроза', <i>челээш</i> 'радуга', <i>хат</i> 'ветер', <i>диңмирээшкин</i> 'гром', <i>шуурган</i> 'буря'	
	Дээр объектилери / Небесные тела / Heavenly bodies	<i>ай</i> 'месяц, луна', <i>хүн</i> 'солнце', <i>сылдыс</i> 'звезда', <i>Шолбан</i> 'Венера', <i>Үгер</i> 'Плеяды', <i>Чеди-хаан</i> 'Большая Медведица'	
	Ландшафт / Landscape	<i>аяң</i> 'горный луг', <i>даг</i> 'гора', <i>хем</i> 'река', <i>хову</i> 'степь', <i>чоога</i> 'ложбина, впадина, овраг', <i>кырлаң</i> 'небольшой горный хребет, отрог', <i>хая</i> 'скала', <i>баалык</i> 'седловина горы'	

Подкласс «Имена собственные» подразделяются на: «человек» — имя, отчество, фамилии, название рода; «животное» — лошадь, корова, собака; «природные объекты» (названия местностей) — топонимы, гидронимы. См. таб. 2.

Таблица 2. Имена существительные в тувинском языке. Имена собственные.

Table 2. Nouns in Tuvan language. Proper names.

Кижиги / Человек / Human	Ат / Имя / Name	<i>Чечек-оол, Артыш, Менди, Кара-кыс</i>
	Адазының ады / Отчество / Middle name	<i>Дүрген-оолович, Бай-Караевна</i>
	Фамилиялар / Фамилии / Surname	<i>Сарыг-оол, Шыырап, Сагаачы</i>
	Аймак-сөөк ады / Названия родов / Names of genera	<i>Кыргыс, Монгуш, Куулар, Түлүш</i>



Дириг амытан / Животное / Animal	Аът /Лошадь / Horse	Калчан-Шилги, Сарала, Доругдай.
	Инек /Корова / Cow	Дөңгүр, Дагыр-Мыйыс, Шокар
	Ыт /Собака / Dog	Ак-Төш, Көстүк, Калдарак
Черлер / Местность / Terrain	Черлер аттары / Топонимы / Placenames	Кызыл, Чаа-Хөл, Кунгуртуз, Бай-Тайга
	Суглар аттары / Гидронимы / Hydronyms	Улуг-Хем, Дус-Хөл, Шивилиг

Непредметные имена (абстрактные понятия) подразделяются на: эмоции, чувственное восприятие, универсальные представления. Семантические пометы и распределение непредметных имен существительных приводится в таблице 3.

Таблица 3. Имена существительные тувинского языка. Непредметные имена.

Tabl. 3. Nouns in the Tuvan language. Non-object names

Туугай билишикиннер / Абстрактные понятия / Abstract concepts	Сагыш-сеткил илерээшкени / Эмоция / Emotion	өөрүшкү 'радость', муңгарал 'горе', дадагалзал 'сомнение'
	Миннишкин / Чувственное восприятие / Sensory perception	дааш 'шум', амдан 'вкус', өң 'цвет', ыт 'звук'
	Ништи билишикиннер / Универсальные представления / Universal submission	болуушкун 'событие', кылдыныг 'действие', байдал 'обстоятельство', үе 'время'

Имена прилагательные подразделяются на классы:

I. Качественные имена прилагательные. В свою очередь, качественные имена подразделяются на подклассы:

- 1) качественные имена прилагательные, характеризующие человека;
- 2) качественные имена прилагательные, характеризующие животное;
- 3) качественные имена прилагательные, характеризующие предмет;
- 4) качественные имена прилагательные, характеризующие природные объекты и явления;
- 5) качественные имена прилагательные, характеризующие абстрактные понятия;

II. Относительные имена прилагательные. Они подразделяются на подклассы:

- 1) относительные имена прилагательных, характеризующие человека;
- 2) относительные имена прилагательных, характеризующие животных;
- 3) относительные имена прилагательных, характеризующие предмет;
- 4) относительные имена прилагательных, характеризующие явления и объекты природы;



5) относительные имена прилагательные, характеризующие абстрактные понятия.

Пример распределения имен прилагательных, характеризующих человека, приведен в таблице 4.

Таблица 4. Имена прилагательные, характеризующие человека.

Tabl. 4. The adjectives characterizing a person.

Кижы / Человек / Human		
Мага-боттуң шынарлары / Физические качества / Physical quality	Мага-бот / Тело, телосложение / Body	моге 'большой и сильный', семис 'полный, толстый', тырың 'крепкий, плотный', эът-ханныг 'здоровый, в теле', ээлгир 'гибкий', ыспан/ыспагар 'худой, тощий'
	Дурт-сын / Рост / Growth	бедик 'высокий', биче 'маленький', чавыс 'низкий', чавыссымаар 'низенький', чолдак 'невысокий, короткий'
	Даштыкы хевир / Внешность / Appearance	арыг-силиг 'чистоплотный', силиг 'аккуратный', чараш 'красивый', шевергин 'с правильными чертами лица', хөрлүг 'видный, представительный'
	Баш / Голова / Head	течик / тецигир 'с выпуклым затылком', дөңгүр 'лысый', тас 'лысый', моң 'с большой головой'
	Баш дүгү / Волосы / Hair	өгбегер/агбагар 'растрепанный', дыдыраш 'кудрявый', суук 'гладкий'
	Арын / Лицо / Face	дырышкак 'морщинистый', хылбаң 'худой, бледный', ыжык 'опухший, распухший', дүгдүнчек 'угрюмый', додуккан 'смуглый'
	Карак / Глаза / Eyes	хапыгыр 'распухший, опухший', бүлдегер 'большой, блестящий', удумзургай 'сонный'



	Кирбик / Брови / Eyebrows	чинге ‘тонкий’, терен ‘густой’, дугаланчак ‘дугообразный’
	Кулак / Уши / Ears	делбиң /делбигир ‘растопыренный’, улуг ‘большой’
	Думчук / Нос / Nose	коң / коңзагар ‘нос крючком’, кырлаң ‘прямой, правильной формы’
	Чаак / Щеки / Cheeks	додуккан ‘смуглый’, бопугур ‘круглощекий’, хорлаңгы ‘обветрившийся’
	Эрин / Губы / Lips	дөрбегер ‘большой’, чингежек ‘тонкий’
	Хавак / Лоб / Forehead	кадыр ‘крутой, высокий’
	Хол / Руки / Hand	хаварык ‘мозолистый, натёртый’, хорлаңгы ‘обветрившийся’
	Оорга / Спина / Back	бүшкүгүр ‘горбатый’, доңгагар ‘сутулый’, ыргак ‘сутулый’, хөкпек/хөкпегер ‘сутулый’.
	Бут / Ноги / Foot	дойтуксумаар ‘прихрамывающий’, майышкак ‘косолапый’, майтак ‘косолапый’
Угаан-медерел талазы-биле шынарлары / Умственные качества/ Mental quality	Эки шынарлар / Положительные Качества / Positive qualities	баштыг ‘умный’, угаанныг ‘умный’, бижик-биликтиг ‘грамотный’, бодангыр-сагынгыр ‘сообразительный’, сагынгыр ‘сообразительный, смышленный’, сарыылдыг ‘разумный’
	Багай шынарлар / Отрицательные качества / Negative qualities эдергейзимээр ‘глуповатый’	мелегей ‘глупый’, сээден ‘тупой’, тудуу ‘слабоумный’, үзээргей ‘глупый’, ээдергей ‘глупый’



Сагъш шынарлары / Психические качества и характер/ Mental qualities and character	Эки шынарлар / Положительные качества / Positive qualities	дидим ‘смелый’, биликсээчел ‘любопытный’, болгаамчалыг ‘острожный’, дөспес ‘непокойный’, дузааргак ‘отзывчивый’, бүзүрээчел ‘доверчивый’, хайгаараачал ‘наблюдательный’, внимательный’
	Багай шынарлар / Отрицательные качества / Negative qualities	адыргак ‘тщеславый’, былдаачал ‘увеливающий, уклоняющийся’, турааргак ‘высокомерный’, туразында ‘своевольный’, туралыг ‘своенравный’, турамык’ ‘высокомерный’, турааргак ‘высокомерный’, хараадаачал ‘сожалеющий’, хедер ‘грубый’, хирээннээчел ‘сердитый, надутый’
Сеткил шынарлары / Душевные Качества / Mental qualitys	Эки шынарлар / Положительные Качества / Positive qualities	ажык ‘доброжелательный’, чазык приветливый, открытый’, амыр-мендилиг ‘приветливый’, арга- сүмелиг ‘участливый’, ажаанзырак ‘заботливый’ баштак ‘веселый, любящий пошутить’
	Багай шынарлар / Отрицательные качества / Negative qualities	кортук ‘трусливый’, дерзии ‘жестокий’, хедер ‘упрямый’, чалгаа ‘ленивый’, халамыргай ‘вялый, апатичный, слабый’, аажылыг ‘дерзкий’, арга-мегелиг ‘хитроумный’, бак ‘плохой’, ылчың ‘несерьезный, легкомысленный, игривый, кокетливый’, шыпыраңнааш ‘вертлявый’



<p>Мөзү-шынар / Нравственные Качества / Moral quality</p>	<p>Эки шынарлар / Положительные качества / Positive qualities</p>	<p>шынчы ‘честный’, томаанныг ‘смирный’, топтут ‘порядочный’, топтут- томаанныг ‘порядочный’ төлептиг ‘достойный’, чазыылдыг ‘выдержанный’, дисциплинированный’, хумагалыг ‘бережливый’</p>
	<p>Багай шынарлар / Отрицательные качества / Negative qualities</p>	<p>мегечи ‘нечестный’, тоожок ‘безнадёжный’, байбац ‘болтливый’, байыыргак ‘кичающийся своим богатством’, балалыг ‘причиняющий вред’, бараа ‘странный’, бачым ‘спешный’, бачыттыг ‘грешный’, буруулуг ‘виноватый’, ёзуургак ‘манерный’, ёозажок ‘неразумный’</p>
<p>Ниитилелде шынарлары / Социальные Качества / Social qualities</p>	<p>Хар-назынының аайы-биле шынарлар / По возрастному признаку / Age symptom</p>	<p>чаш ‘младенческий’, чалыы ‘молодой, юный’, ушпа ‘дряхлый, престарелый’, хеймер ‘самый младший в семье’, хензиг ‘маленький’, хоочун ‘старый’, элээди ‘подросткового возраста’, ылбыс ‘новорождённый’, чөнүк ‘дряхлый, престарелый’, чедишкен ‘достигший сорока пяти- пятидесяти лет’.</p>
	<p>Өнчү-хөрөңги аайы- биле шынарлар / По имущественному признаку / By ownership</p>	<p>бай ‘богатый’, тодуг ‘сытый, богатый’, тодуг- догаа ‘зажиточный, богатый, обеспеченный’, түрөңги ‘нищий’, чединмес ‘нуждающийся’, ядыы ‘бедный’, ядамык ‘бедный’</p>



	<p>Ниитилелде, өг-бүледе кижиниң туружунуң аайы-биле шынарлар / Признание / не признание в обществе, семье, коллективе/ Recognition / recognition in society, the family, the community</p>	<p>барыктыг ‘сносный’, үлегерлиг ‘образцовый, примерный’, чааскаан ‘одинокый’, чаңгыс ‘одинокый’, хүндүлүг ‘уважаемый’, хүндүткелдиг ‘почётный, авторитетный’, хүндүтен ‘почтенный’, ынак ‘любимый’, эп-найыралдыг ‘дружный, спаянный’, эп-сеткилдиг ‘единодушный, спаянный, дружный’</p>
<p>Салым-чаяан / Способности, таланты / Abilities, talents</p>		<p>салымныг ‘способный, одарённый, талантливый’, шевер ‘искусный, умелый’, чаяанныг ‘способный, одарённый, талантливый’</p>

Базы данных лексического фонда тувинского языка служат также для выявления лексической сочетаемости лексем. В автоматизированной системе содержится информация о семантической сочетаемости имен прилагательных и имен существительных; наречий и глаголов; имен существительных и глаголов, а также о невозможности семантически недопустимых сочетаний. Так, например, автоматизированной системой будут исключаются семантически недопустимые сочетания, например: аяар самбыра ‘тихая доска’. Таким образом, создание разных пользовательских запросов с учетом семантики позволяет уточнить, выявить правила сочетаемости тех или лексических единиц.

Над проектом работают научные сотрудники НОЦ «Тюркология» ТувГУ — языковеды-тюркологи: к. филол. н. Б. Ч. Ооржак (руководитель проекта), к.филол. н. А. Б. Хертек; ст. преподаватель кафедры тувинской филологии и общего языкознания, к. филол. н. А. Я. Салчак, доцент кафедры русского языка и литературы, к. филол. н. В. С. Ондар ; писатель, переводчик, заведующий литературным музеем ТувГУ М. А. Кужугет ; литератор, доцент кафедры тувинской филологии и общего языкознания, к. филол. н. Е. Т. Чамзырын. Разработку электронных программ производит доцент кафедры информатики, к. физ.-мат. н. С. М. Далаа.

Заключение

Создаваемая база данных лексического фонда тувинского языка и ее семантическая разметка как электронная справочно-поисковая система призваны обеспечить автоматизацию работы пользователей с большими объемами текстов на тувинском языке. Это повысит эффективность работы ЭКТТЯ для поиска необходимых фрагментов текстов с искомой семантической информацией. Электронная система для работы с текстом поднимет на новый уровень работу по разработке учебно-методической базы по преподаванию тувинского языка в школе и вузе, подготовку разного типа учебников, а также исследования по тувинскому языку и литературе и составление словарей, начиная от учебных до академических.

Первые результаты работы над разработкой семантической разметки электронного корпуса тувинского языка уже апробировались в докладах на конференциях: III Международной конференции по компьютерной обработке тюркских языков «TurkLang 2015» (Казань, 17–19 сентября 2015); Международной конференции по компьютерной и когнитивной лингвистике TEL-2016 (Казань, 21–24 апреля 2016); Всероссийской научной конференции «Языки народов Сибири и сопредельных регионов» (Новосибирск, 11–14 октября 2016). Опубликованы статьи (Ооржак, Хертек, 2015; Создание базы ..., 2016).

СПИСОК ЛИТЕРАТУРЫ

Бавуу-Сюрюн, М. В., Далаа, С. М. Морфемно-орфографический словарь тувинского языка [Электронный ресурс] // Электронный корпус текстов тувинского языка. URL: <http://www.tuvacorpus.ru/?q=content/slovari> (дата обращения: 12.09.2016).

Ооржак, Б. Ч., Хертек, А. Б. (2015) Разработка семантической разметки электронного корпуса тувинского языка // Материалы 3-й Международной конференции по компьютерной обработке тюркских языков «TurkLang 2015». Казань, 17–19 сентября 2015. Казань : Изд-во АН Республики Татарстан. С. 351–362.

Создание базы данных лексического фонда тувинского языка (2016) / Ооржак, Б. Ч., Хертек, А. Б., Кужугет, М. А., Салчак, А. Я., Ондар, В. С., Чамзырын, Е. Т. // Труды Международной конференции по компьютерной и когнитивной лингвистике. TEL-2016. Казань, 21–24 апреля 2016. Казань : Изд-во Казанского государственного университета. Вып. 17. 392 с. С. 278–281.

Дата поступления: 15.10.2016 г.

REFERENCES

Bavuu-Siuriun, M. V. and Dalaa, S. M. Morfemno-orfograficheskii slovar' tuvinskogo iazyka. *Elektronnyi korpus tekstov tuvinskogo iazyka* [online] Available at: <http://www.tuvacorporus.ru/?q=content/slovari> (access data: 12.09.2016). (In Russ.).

Oorzhak, B. Ch. and Khertek, A. B. (2015) Razrabotka semanticheskoi razmetki elektronnoho korpusa tuvinskogo iazyka. In: *Materialy 3-ei Mezhdunarodnoi konferentsii po komp'iuternoi obrabotke tiurkskikh iazykov «TurkLang 2015»*. Kazan', 17–19 sentiabria 2015. Kazan', Izd-vo AN Respubliki Tatarstan. Pp. 351–362. (In Russ.).

Cozdanie bazy dannykh leksicheskogo fonda tuvinskogo iazyka (2016) / Oorzhak, B. Ch, Khertek, A. B., Kuzhuget, M. A., Salchak, A. Ia., Ondar, V. S. and Chamzyryn, E.T. In: *Trudy Mezhdunarodnoi konferentsii po komp'iuternoi i kognitivnoi lingvistike. TEL-2016. Kazan', 21–24 apreliia 2016*. Kazan', Izd-vo Kazanskogo gosuniversiteta. Vol. 17. 392 p. Pp. 278–281. (In Russ.).

Submission data: 15.10.2016.

Библиографическое описание статьи:

Ооржак Б. Ч., Хертек А. Б., Кужугет М. А., Ондар В. С. Семантическая разметка имени для Электронного корпуса текстов тувинского языка [Электронный ресурс] // Новые исследования Тувы. 2016, № 4. URL: <http://nit.tuva.asia/nit/article/view/615> (дата обращения: дд.мм.гг.).

Citation:

Oorzhak B. Ch., Khertek A. B., Kuzhuget M. A., Ondar V. S. Semantic markup of nouns and adjectives for the Electronic corpus of texts in Tuvan language. *Novye issledovaniia Tuvy*, 2016, no. 4 [on-line] Available at: <http://nit.tuva.asia/nit/article/view/615> (accessed: ...).