

УДК 811.512.154:81-13

КОРПУСНОЕ ИССЛЕДОВАНИЕ ПРИЛАГАТЕЛЬНОГО “ЧОҢ” В КЫРГЫЗСКОМ ЯЗЫКЕ

¹Чортекова М.Т., ²Найманова Ч.К.

¹*Кыргызский государственный университет имени И. Арабаева, Бишкек,
e-mail: mhortekova@gmail.com;*

²*Кыргызско-Турецкий университет «Манас», Бишкек, e-mail: cholpon.naymanova@manas.edu.kg*

Настоящая статья анализирует прилагательное “чоң” (большой, великий) в контексте корпуса кыргызского языка. Актуальность исследования определяется отсутствием исследований корпуса кыргызского языка. Изучаются теоретические предпосылки исследования прикладного языкознания, компьютерной лингвистики, автоматизация обработки естественного языка и корпусной лингвистики. Лингвистический корпус кыргызского языка был внедрен в 2020 г. при финансировании Deutscher Akademischer Austauschdienst (DAAD) при участии Кыргызско-Турецкого университета «Манас» (Кыргызская Республика) и университета Саарланда (Германия). Корпус состоит из двух частей, первая часть вмещает 1 205 888 слов 84 художественных текстов пяти жанров: из романов, повестей, эпосов, малых эпосов и сказок. Корпус аннотирован тегами леммы и частей речи, а также богатыми потекстовыми метаданными. Тексты предоставлены фондом “Биздин мурас”, который способствует развитию кыргызского языка. Во второй части добавлены кыргызские пословицы (также из базы текстов фонда “Биздин мурас”) и около 1 млн слов газетного текста предоставлено государственной официальной газетой Кыргызской Республики “Эркин-Тоо”. Объектом исследования выступает прилагательное “чоң” в контексте корпуса кыргызского языка. Проведен анализ корпусной репрезентации прилагательного “чоң” в разных текстах. В статье используются описательно-статистический и лексико-семантический анализ прилагательного “чоң”. Примеры прилагательных получены методом сплошной выборки.

Ключевые слова: имя прилагательное, чоң (большой, великий), кыргызский язык, компьютерная лингвистика, корпусная лингвистика, теги, корпус

CORPUS STUDY OF THE ADJECTIVE “CHONG” IN KYRGYZ LANGUAGE

¹Chortekova M.T., ²Naymanova Ch.K.

¹*I. Arabaev Kyrgyz State University, Bishkek, e-mail: mhortekova@gmail.com;*

²*Kyrgyz-Turkish University “Manas”, Bishkek, e-mail: cholpon.naymanova@manas.edu.kg*

This article analyzes the adjective “chong” (“big”) in the context of the corpora of the Kyrgyz language. The relevance of research is determined by the lack of corpora research of the Kyrgyz language. The literature review in applied linguistics, computer linguistics, natural language processing and corpus linguistics is studied. The linguistic corpus of the Kyrgyz language was introduced in 2020 under the financial support of the Deutscher Akademischer Austauschdienst (DAAD). The linguistic corpus is done by the scholars of the Kyrgyz-Turkish University “Manas” (Kyrgyz Republic) and the scholars of the University of Saarland (Germany). The corpus consists of two parts, the first part includes 1,205,888 words of 84 literary texts of five genres: novels, short stories, epics, small epics and fairy tales. The corpus is annotated with lemma tags and parts of speech tags, as well as rich text metadata. The texts were provided by the “Bizdin Muras” foundation, that promotes the development of the Kyrgyz language. In the second part of the corpus Kyrgyz proverbs are added; about 1 million words from newspaper texts were provided by the “Erkin-Too” state official newspaper of the Kyrgyz Republic “Erkin-Too”. The object of research is the adjective “chong” (“big”) in the context of the corpus of the Kyrgyz language. Corpus representation of the adjective “chong” (“big”) in different texts was carried out. The article uses descriptive-statistical and lexical-semantic analysis of the adjective “chong” (“big”). Examples of adjectives obtained by the continuous sampling method.

Keywords: adjective, “chong” (“big”), Kyrgyz language, computer linguistics, corpus linguistics, tags, corpus

В начале XXI в. лингвистическая научная парадигма стала междисциплинарной, прикладные аспекты вышли на первый план при изучении лингвистических фактов. Во второй половине XX в. появилась автоматизация обработки естественного языка (Natural Language Processing, NLP). При выполнении задач NLP компьютер читает текст, проверяет его на правильность, выполняет разные инструкции по значению текста. Для углубленного понимания вопросов человеческого общения и автома-

тизации естественного языка лингвистика опирается на компьютерную лингвистику (Computational Linguistics, CL) [1].

Цель исследования – анализ прилагательного “чоң” в контексте корпуса кыргызского языка.

Материалы исследования составляют примеры из лингвистического корпуса кыргызского языка, методами исследования выступают описательно-статистический и лексико-семантический анализ прилагательного “чоң”.

Теоретические основы исследования. Компьютерная лингвистика занимается компьютерным моделированием владения естественным языком и решением прикладных задач автоматической обработки текстов и речи. Данная отрасль берет свое начало в 1950-х гг. из исследований американского лингвиста, основоположника генеративной грамматики Ноама Хомского по структуризации естественного языка, а также из экспериментов машинного перевода. В 1954 г. в Джорджтаунском университете (США) ученые провели первый в мире открытый эксперимент по машинному переводу; тогда ученые выполнили более 60 предложений текстов о космосе с русского языка на английский. Компьютерная лингвистика использует компьютерные средства для обработки языковой информации. В ее основе лежат теории искусственного интеллекта, антропологии и лингвистики; таким образом, задачами компьютерной лингвистики являются:

а) создание инструментов, которые могут извлекать лингвистическую информацию из текстов;

б) распознавание лингвистических признаков в текстах;

в) создание ресурсов для хранения лингвистических единиц, т.е. создание лингвистического корпуса [1].

Компьютерная лингвистика реализуется в следующих платформах:

а) машинный перевод, который использует искусственный интеллект для перевода текстов с исходного языка на другой. Google Translate является хорошим примером машинного перевода;

б) чат-боты: программы, имитирующие человеческий разговор посредством устной или письменной речи, обычно для обслуживания клиентов. Во многих крупных компаниях мира наряду с телефоном и электронной почтой доступен и чат программы;

в) извлечение информации: создание данных из неструктурированных и структурированных текстовых источников. Примером может служить Википедия, продукт случайных редакторов, который можно использовать для тренировки точности и полноты извлечения открытой информации;

г) естественно-языковой пользовательский интерфейс: инструменты такого типа позволяют людям взаимодействовать с операционными системами наших устройств с помощью прикладного диалога. Например, Siri, облачный персональный помощник и вопросно-ответная система, программный клиент которой входит в состав iOS, iPadOS, watchOS, macOS и tvOS компании Apple; также Alexa – виртуальный помощник

Amazon, который в основном конкурирует с Google Assistant и Apple Siri;

д) анализ настроений – анализ обработки естественного языка, при котором обнаруживается положительное или отрицательное настроение отрывка текста [1].

Корпусная лингвистика. Корпусная лингвистика изучает язык в его текстовом корпусе. Лингвистические данные собираются в естественном контексте этого языка – с минимальным экспериментальным вмешательством. Языковой корпус, или корпус текстов, – унифицированная, структурированная, филологически компетентная база языковых данных, предназначенная для решения определенных лингвистических задач. Основными признаками современного лингвистического корпуса являются машиночитаемый формат, репрезентативность, наличие металингвистической информации [2].

Метод текстового корпуса использует совокупность текстов, написанных на любом естественном языке, для получения набора абстрактных правил, которые управляют этим языком. Эти результаты можно использовать для изучения отношений между этим предметным языком и другими языками, которые подверглись аналогичному анализу. Первые такие корпусы составлялись вручную из исходных текстов, но теперь эта работа автоматизирована [3].

Корпусы использовались не только для лингвистических исследований, они также использовались для составления словарей, таких как The American Heritage Dictionary of the English Language («Словарь американского наследия английского языка», 1969), и руководств по грамматике, таких как A Comprehensive Grammar of the English Language («Всеобъемлющая грамматика английского языка», 1985).

Первый современный лингвистический корпус стандартного американского английского языка – Брауновский корпус (“The Brown Corpus of Standard American English”) был разработан в Брауновском университете (США) учеными Н. Френсисом (Nelson Francis) и Г. Кучера (Henry Kucera) в 1961 г. Он состоит из 1 млн слов из текстов американского английского языка [4].

Чтобы корпус соответствовал требованиям хорошего стандартного справочника, тексты были отобраны в разных пропорциях из 15 различных текстовых категорий: пресса (репортажи, редакционные статьи, обзоры), навыки и хобби, религиозные тексты, познавательные/научные тексты, художественная литература (различные подкатегории) и т.д.

Сегодня этот корпус, хотя считается небольшим и слегка устаревшим, все еще используется. Макет Брауновского корпуса был скопирован другими разработчиками корпуса. Корпус британского английского (LOB, Lancaster-Oslo-Bergen) и корпуса индийского английского языка (Kolharur Corpus) являются двумя примерами корпусов, созданных на основе Брауновского корпуса. Оба они состоят из 1 млн слов письменного языка (500 текстов по 2000 слов каждый), отобранных из тех же 15 категорий, что и Брауновский корпус. Долгое время Брауновский корпус и корпус LOB были единственными доступными лингвистическими корпусами. Поэтому многие исследования в области корпусной лингвистики были основаны на этих корпусах.

Еще одним важным «малым» корпусом является Лондонско-Лундский корпус разговорного британского английского (LLC), который был разработан в Лундском университете под руководством Дж. Свартвика. Это первый компьютерно читаемый корпус разговорной речи и состоит из 100 разговорных текстов около 5000 слов каждый. Тексты классифицируются по разным категориям, таким как спонтанная беседа, спонтанный комментарий, спонтанная и подготовленная речь и т.д. Тексты орфографически расшифрованы и снабжены подробной просодической разметкой.

С 1980 г. группа ученых начала работать над составлением словаря Collins Cobuild English Language Dictionary, они собирали корпус текстов на компьютере для создания словарей и изучения языка и имели ежедневный доступ к корпусу примерно из 20 млн слов. В корпус были добавлены новые тексты, и в 1991 г. он был запущен как Bank of English (BoE). В корпус Bank of English (BoE) добавляется все больше и больше данных, а последний выпуск (1996 г.) содержит около 320 млн слов [4].

В корпус постоянно добавляются новые материалы, чтобы он «отражал современное состояние современного английского языка». Корпус такого рода, который с помощью новых дополнений «следит» за изменениями в языке, называется мониторингом корпусом.

В 1995 г. был выпущен еще один большой корпус – Британский национальный корпус (British National Corpus, BNC). Этот корпус состоит примерно из 100 млн слов. Как и Bank of English (BoE), он содержит как письменный, так и устный материал, но, в отличие от Bank of English (BoE), (British National Corpus, BNC) является конечным – после его завершения к нему больше не добавляются тексты [4].

Тексты BNC были отобраны в соответствии с заранее определенными критериями отбора с установленными целевыми значениями количества текста, которое должно быть включено из разных типов текста. Тексты закодированы с разметкой, предоставляющей информацию о текстах, авторах, докладчиках.

По состоянию на ноябрь 2021 г. Корпус современного американского английского (СОСА, 2008) состоит из 1 млрд слов. Корпус постоянно растет: в 2009 г. он содержал более 385 млн слов; в 2010 г. корпус вырос до 400 млн слов; к марту 2019 г. корпус вырос до 560 млн слов. По состоянию на ноябрь 2021 г. Корпус современного американского английского языка состоит из 485 202 текстов. Согласно веб-сайту корпуса, текущий корпус (ноябрь 2021 г.) состоит из текстов, включающих 24–25 млн слов за каждый год с 1990 по 2019 [4].

Для каждого года (1990–2019) корпус равномерно разделен между шестью регистрами/жанрами: телевидение/фильмы, устная речь, художественная литература, журнал, газета и академический дискурс.

В дополнение к шести регистрам, которые были перечислены ранее, СОСА (по состоянию на ноябрь 2021 г.) также содержит 125 496 215 слов из блогов и 129 899 426 слов с веб-сайтов, что делает его корпусом, действительно отражающим состояние современного американского английского языка [4].

Корпусный анализ лингвистических единиц проводится с целью выявления определенных правил использования языка, например, грамматических или лексических моделей, которые относятся к определенному жанру или типу текста, которые являются ценным источником для лингвистического исследования.

Официальный лингвистический корпус кыргызского языка не был внедрен до 2020 г. В том году при финансировании Deutscher Akademischer Austauschdienst (DAAD) доцент кафедры перевода Кыргызско-турецкого университета “Манас” (Кыргызская Республика) Аида Касиева и ученые университета Саарланда (Германия) Jörg Knappen, Stefan Fischer и Elke Teich начали разработку аннотированного корпуса кыргызского языка, который состоит из двух частей. Первая часть вмещает 1 205 888 слов 84 художественных текстов пяти жанров: из романов, повестей, эпосов, малых эпосов и сказок. Корпус аннотирован тегами леммы и частей речи, а также богатыми потекстовыми метаданными. Тексты предоставлены фондом “Биздин мурас”, который способствует развитию кыргызского языка.

Во второй части добавлены кыргызские пословицы (также из базы текстов фонда “Биздин мурас”) и около 1 млн слов газетного текста предоставлено государственной официальной газетой Кыргызской Республики “Эркин-Тоо” [5].

В данной научной статье делается попытка корпусного анализа прилагательного “чоң” в кыргызском языке. Прилагательное “чоң” имеет следующие значения: 1) большой, огромный, великий; 2) старший по возрасту; 3) начальник, чиновник.

В корпусе прилагательное “чоң” используется 1858 раз в 240 различных текстах (в 2 493 894 словах [1019 текстов]; частота: 745 020 примеров на миллион слов). В эпосе “Манас” прилагательное “чоң” используется 622 раз в корневой морфеме [6].

Данный лингвистический корпус также собрал данные прилагательного “чоң” из других малых эпосов (“Олжобай менен Кишимжан”, “Кожожаш”, “Эр Тоштук”), народных сказок и литературных произведений “Джамила”, “Тоолор кулаганда” (“Когда падают горы”) великого кыргызского писателя Ч. Айтматова, а также из романа “Келкел” кыргызского писателя Т. Касымбекова.

В эпосе “Манас” прилагательное “чоң” используется часто в словосочетаниях *чоң казат, чоң уруш, чоң согуш, чоң жаңжал, чоң кыргын, чоң чыр (великий поход, великая схватка, великая война, великое сражение)*. По сути эпос “Манас” является героическим эпосом, его сюжет в основном развивается вокруг войн за свободу и единство кыргызского народа во главе с героем Манасом. Например: *Көк жал Манас барында Кытайга салган чоң жаңжал (Великое сражение против Китая при жизни героя Манаса). Кыраан Манас барында Кылымга салгын чоң жаңжал (Великая война века при жизни героя Манаса)*.

Под словосочетанием *Чоң Бээжин (большой Пекин)* подразумевается столица вражеских племен и народов. Данное словосочетание также многократно используется в эпосе “Манас”. Например: *Чоң Бээжин толкуп, көп кол каптады (Взбунтовался большой Пекин, и в него вступила многочисленная армия)*.

Герой Манас и в детстве от своих сверстников отличался храбростью, лидерскими качествами и прозорливостью. Его родители боялись, что злые духи сглазят его, поскольку он был долгожданным единственным ребенком в семье, и называли его прозвищем “*Чоң жинди*” (*Великий сумасброд*). Например: *Уулунун аты Чоң Жинди (Его сына зовут “Чоң Жинди” (Великий сумасброд). Онго жашың чыккыча Кара калмак,*

манжууга Чоң Жинди – деп ат койдук (Мы назвали тебя Чоң Жинди до твоего десятилетия, чтобы калмыки, манжу звали тебя по этому имени). Наряду с этим прозвищем у него было свое имя Манас, которое также использовалось вместе с прилагательным “чоң”. Например: *Чокудагы чоң Манас чогула жандан түңүлдү (Великий, храбрый Манас потерял надежду)*.

Прилагательное “чоң” также ассоциируется с величиной, объемом, величавостью крепостей, географических местностей: рек, гор, холмов, озер и др. Например: *Бийик соккон чоң себил (высокая, огромная крепость), Чоң Алтайдын туурасы (широта величественного Алтая), Чоң Илени кыдырып (бродя по большой долине Иле), Чоң өзөндүн боюна (на берегу большой реки), Чоң Таласка токтоду (остановился в величественном Таласе), Чоң Ак – Чийдин боюнда (на берегу крупной реки Ак-Чий), Чоң Букардын айылында (в ауле величественной Бухары), Чоң Кең – Колдун боюнда (на берегу широкой реки Кең-Кол), чоң чынар терек (огромный тополь), чоң жол (широкая дорога), чоң кара таи (громадный камень)*,

Эпос также прилагательным “чоң” описывает предметы: *чоң сыр кесе (большая эмалированная чашка), чоң казан (большой казан), чоң челек (большое ведро), чоң өргөө (большой дом), чоң дарбаза (высокие ворота), чоң сөйкө (большие красивые серьги), чоң чепкен (громадный плащ), чоң байге (большой приз), чоң камчы (большой хлыст), чоң дүрбү (большой бинокль), чоң чокой (широкая обувь), чоң килем (длинный ковер)*.

Абстрактные понятия также выражаются прилагательным *чоң*: *чоң тамаша-жыргал (большое веселье, радость), чоң жини (сильная злость), чоң күч (огромная сила), чоң кек (глубокая месть), чоң бакты (большое счастье), чоң зыян (большие потери), чоң бүлүк (большой раздор)*.

У кыргызов понятие бог выражалось лексемой *кудай*, однако использование прилагательного *чоң* для данного слова передает могущество, сакральность бога (*чоң кудай*). Также кыргызы выражали почтение своим кочевым правителям и признавали превосходство их власти, используя лексему *чоң*. Например: *чоң баша (великий правитель), чоң так (могучий трон)*.

Эпос богат стилистическими средствами в использовании лексемы *чоң*. Кыргызы особо ценили лошадей, они боялись, что злые духи сглазят их, использовали эвфемизмы для названия лошадей. Например, метонимия *Чоң күрөң* выражает *бурую лошадь*.

Прилагательное “чоң” также используется в языке исторического романа “Сынган Кылыч” («Сломанный меч») кыргызского писателя Тологона Касымбекова. Данный лингвистический корпус также собрал данные использования прилагательного “чоң” из вышеуказанного произведения. Роман описывает исторические события кыргызского народа во второй половине XIX в. Язык романа отличается изобилием образительно-выразительных средств, в том числе искусное мастерство писателя в использовании языковых единиц делает сюжет романа незабываемым. Кочевые кыргызы вошли в состав Кокандского ханства, которое они считали своим домом, своим гнездом. Например: *Бул чоң уя! Уяны каргалар ээлеп алышкан (Это большое гнездо, в нем обитают «вороны»)*. Сюжет развивается вокруг дворцовых интриг и переворотов: *талашкан чоң уя өлкөнүн ордосу Кокон (столицей большого государства, «гнезда» является Коканд)*.

Прилагательное “чоң” также содействует описанию событий того времени. Например: *чоң базар (огромный базар), чоң шаар (большой город), чоң дарбаза (большие ворота), чоң өргөө (большая юрта), чоң оюн (большие интриги), чоң той (большой пир), чоң өлкөлөр (сильные государства), чоң ханике (великая царица)*. Прилагательное “чоң” часто используется в сочетании с историзмами: *чоң акимдер (влиятельные губернаторы), чоң кара сакалчан сарбаз (высокий чернобородый пеший солдат), чоң бий (влиятельный правитель)*. Таким образом, прилагательное “чоң” иногда интерпретирует *влиятельность*, поскольку современный кыргызский язык использует метафору “чоң” для выражения семантики лексемы «государственный влиятельный чиновник».

Прилагательное “чоң” изображает элементы этнической традиции и менталитета. Например: *чоң чанач (большой кожаный сосуд, в котором готовится и перевозится кумыс), чоң арабакеч (искусный арбакеш, ломовой извозчик), чоң кесе айран (большая чаша кисломолочного напитка), чоң кестик (большой булатный нож)* и др. В современном кыргызском языке прилагательное “чоң” также интерпретирует семантику лексемы «искусный, опытный, профессиональный, квалифицированный». Например: *чоң комузчу (искусный комузист), чоң жазуучу (выдающийся писатель)*.

Лингвистический корпус также показывает, что прилагательное “чоң” используется в сочетании с наименованиями флоры и фауны. Например: *чоң боз ат (большой белый конь), чоң чынар (высокий/большой*

тополь), чоң илбирс (большой барс), чоң ак илбирс (большой снежный барс), чоң жылан (большая змея), чоң алма (большое яблоко), чоң кара эчки (большая черная коза), чоң аюу (большой медведь) и др.

Прилагательное “чоң” также определяет качество мифологических понятий. Например: *чоң Көкдөө, чоң Малгун (Великан)*.

Прилагательное “чоң” также используется в газетных текстах, примеры были собраны методом сплошной выборки. Пословицы в определенной мере отражают менталитет и культуру кыргызского народа, например: *Ден соолук чоң байлык (Здоровье – большое богатство); Ар кимдин тилеги өзүнө чоң (У каждого свои большие желания); Кичине жалкоолук чоң жалкоолук жаратат (Маленькая лень до большой выростет); Кичине жалаадан чоң жалаа чыгат (Маленькая клевета до большой выростет); Корккондун көзү чоң (У труса в глазах двоится); Бербеген кудайга чоң дооң барбы (У вас есть большая жалоба на бога?); Үй – бүлөнү бузган – чоң шаарды бузганга барабар (Разрушение семьи равносильно уничтожению большого города); Чоң иш бүтүрөм десең, майдадан башта (Хочешь закончить большую работу, начни с малого); Чоң кемеге – чоң толкун (Большая волна для большого корабля); Чоң чатактан кичинекей достук артык (Лучше маленькая дружба, чем большая ссора); Ырчылык – чоң өнөр (Пение – великое искусство); Чоң дарыянын акканы жай (Большая река течет тихо); Байды чоң сүйлөткөн, кородогу койлору (Богатый человек кичится количеством своих овец)*.

В газетных текстах данное прилагательное содействует полной передаче прилагательного “чоң”. Например: *Балдар эмес, чоң эле адамдар өздөрү интернет оюндарында сааттап ойношот (Не только дети, но взрослые люди играют в компьютерные игры). Дүйнөлүк интернет оюндары өзүнчө чоң тема (Мир онлайн-игр – отдельная большая тема)*.

В стиле газеты прилагательное имеет нейтральную семантику, например: *Бул багытта дагы да көптөгөн долбоорлор ишке ашырылып, карапайым калк үчүн өтө чоң пайдалуу жумуштар аткарылууда (В этом направлении реализуется еще много проектов и делается много полезной работы для простых людей)*.

Газетные тексты интерпретируют национальную политику государства, проблемы экономики, социальной жизни и культуры народа. Например: *Заводдун адистерин кармап калыш үчүн өтө чоң эмгек талап кылынды (Пришлось немало потрудиться, чтобы удержать специалистов завода)*.

Ушул кыйынчылыктарды эстесем, элге, коллективге чоң кызмат кылган экенмин деп азыр ойлойм (Когда я вспоминаю о тех трудностях, думаю, что немало послужил людям и команде). ЕАЭБге киргенбиз, ал үчүн өтө чоң аракет кылдык (Мы вступили в ЕАЭС, для этого приложили немало усилий). Мамлекетибиздин унаа жол тармагында эмгектенген жолчулардын мээнетти чоң (Сотрудники дорожной отрасли вносят огромный вклад). В вышеуказанных примерах прилагательное “чоң” участвует в создании микроконтекста в рамках предложений и абзацев. Иногда оно выступает в качестве определения в предложении, например: *Өзгөн элине чоң иши – чара уюштуруп берип жүрөт (Он организует грандиозные мероприятия для народа Узгенского района). Марат Акимович келгенден бери биздин институтта жакшы жагына аябай чоң өзгөрүүлөр болду (С приходом Марата Акимовича в нашем институте произошли большие изменения в лучшую сторону). Эженин театрда, кинодо орду чоң эле (У нее были огромные заслуги в театре и кино). Маалыматтарды таратууда ватсап группалары да чоң көмөкчү болду (Группы WhatsApp* сыграли большую роль в распространении информации).*

*Компания Meta Platforms Inc., владеющая WhatsApp, признана экстремистской организацией, ее деятельность на территории России запрещена.

Заклучение

Данная статья проанализировала корпусное использование прилагательного “чоң” в кыргызском языке. Исследования в

таком русле актуальны, когда вопросы прикладной лингвистики выходят на первый план. Корпус зарегистрировал 622 примера прилагательного “чоң” из эпоса “Манас”, 126 примеров из романа «Сломанный меч», из повести “Джамила” 23 примера, из эпоса “Олжобай и Кишимжан” 43 примера, из романа «Когда падают горы» 54 примера, из разных сказок 71 пример, из малого эпоса “Кожожаш” 13 примеров, из эпоса “Эр Тоштук” 110 примеров, из романа “Келкел” 16 примеров, из романа “Кыргын” 94 примера; прилагательное “чоң” также используется в 45 пословицах и поговорках и 254 прилагательных из газетных статей собраны методом сплошной выборки из газеты “Эркин-Тоо”.

Список литературы

1. Schubert L.K. Computational Linguistics. The Stanford Encyclopedia of Philosophy. Stanford, CA. 2020. [Электронный ресурс] URL: <https://plato.stanford.edu/archives/spr2020/entries/computational-linguistics> (дата обращения: 06.07.2023).
2. Gena R. Bennet. Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers // Language Value. 2010. Vol. 2. P. 140–143.
3. Kübler S., Zinsmeister H. Corpus Linguistics and Linguistically Annotated Corpora. Bloomsbury, 2015. 320 p.
4. Corpus linguistics. [Электронный ресурс] URL: https://www1.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/history.html (дата обращения: 26.06.2023).
5. Kasieva A., Knappen J., Fischer S., Teich E. A new Kyrgyz corpus: sampling, compilation, annotation. 2020. [Электронный ресурс] URL: <https://www.zfs.uni-hamburg.de/dgfs2020/programm/abstracts/dgfs2020-clp-kasieva.pdf> (дата обращения: 26.05.2023).
6. Корпус кыргызского языка. [Электронный ресурс]. URL: https://corpora.clarin-d.uni-saarland.de/cqpweb/kyrgyz_2022_10_03/ (дата обращения: 06.07.2023).