

2015

“МАНАС” ЭПОСУНУН ТЕКСТТИК
КОРПУСТАРЫНЫН НЕГИЗИНДЕ ГИПЕРТЕКСТТҮҮ СӨЗДҮКТҮ ТҮЗҮҮ

МЭЭРИМ УЙЧУБЕК КЫЗЫ



КЫРГЫЗ-ТҮРК «МАНАС» УНИВЕРСИТЕТИ

ТАБИГЫЙ ИЛИМДЕР ИНСТИТУТУ

КОМПЬЮТЕР ИНЖЕНЕРДИГИ БӨЛҮМҮ

**“МАНАС” ЭПОСУНУН ТЕКСТТИК
КОРПУСТАРЫНЫН НЕГИЗИНДЕ ГИПЕРТЕКСТТҮҮ
СӨЗДҮКТҮ ТҮЗҮҮ**

Даярдаган

Мээрим УЙЧУБЕК КЫЗЫ

Илимий жетекчи

Доц. Др.Бакыт ШАРШЕМБАЕВ

МАГИСТРДИК ДИССЕРТАЦИЯ

Чын куран 2015

КЫРГЫЗСТАН/БИШКЕК

КЫРГЫЗ-ТҮРК «МАНАС» УНИВЕРСИТЕТИ

ТАБИГЫЙ ИЛИМДЕР ИНСТИТУТУ

КОМПЬЮТЕР ИНЖЕНЕРДИГИ БӨЛҮМҮ

**“МАНАС” ЭПОСУНУН ТЕКСТИК
КОРПУСТАРЫНЫН НЕГИЗИНДЕ ГИПЕРТЕКСТТҮҮ
СӨЗДҮКТҮ ТҮЗҮҮ**

(МАГИСТРДИК ДИССЕРТАЦИЯ)

Мээрим УЙЧУБЕК КЫЗЫ

Илимий жетекчи

Доц. Др. Бакыт ШАРШЕМБАЕВ

Чын куран 2015

КЫРГЫЗСТАН/БИШКЕК

İNTİHAL YAPILMADIĞINI BELİRTEN İFADE

Ben bu tezdeki bütün bilgileri akademik ve etik kurallarına göre aldığımı ve sunduğumu belirtiyorum. Bu çalışmaya özgün olmadan kullandığım bütün materyal ve bilgilere akademik ve etik kurallar gereğince atıfta bulunduğumu ve hiçbir şekilde intihal yapmadığımı belirtiyorum.

İSİM, SOYAD: **Meerim UYÇUBEK KIZI**

İMZA:

TARİH:

ИЛИМИЙ ЭТИКАГА ЫЛАЙЫКТУУЛУК ТУУРАЛУУ БИЛДИРҮҮ

Мен бул эмгекте алынган бардык маалыматтарды академиялык жана этикалык эрежелерге ылайык колдондум. Тагыраак айтканда бул эмгекте колдонулган бирок мага тиешелүү болбогон маалыматтардын бардыгын колдонулган адабияттар бөлүмүндө көрсөттүм жана эч кайсы жерден плагиат жасалбагандыгына ынандырып кетким келет.

АТЫ, ЖӨНҮ: **Мээрим УЙЧУБЕК КЫЗЫ**

КОЛУ:

ДАТАСЫ:

YÖNERGEYE UYGUNLUK

“Manas Destanındaki Metinlerin Konutlarına Göre Hypertextlik Sözcük Düzenlemek”adlı Yüksek Lisans tezi, Kırgızistan Türkiye Manas Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan:

Meerim UYÇUBEK KIZI

İmza:

Danışman:

Doç.Dr. Bakıt ŞARŞEMBAEV

İmza:

Bigisayar Mühendisliği ABD Başkanı

Doç.Dr. Rayimbek SULTANOV

İmza:

KABUL VE ONAY

Doç.Dr. Bakıt ŞARŞEMBAYEV'in danışmanlığında Meerim UYÇUBEK KIZI tarafından hazırlanan ***“Manas Destanındaki Metinlerin Konutlarına Göre Hypertextlik Sözcük Düzenlemek”*** adlı bu çalışma, jürimiz tarafından Kırgızistan Türkiye Manas Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

04.06.2015

JÜRİ:

Danışman Doç. Dr. Bakıt ŞARŞEMBAYEV

Üye Doç. Dr. Sabira NURZHANOVA

Üye Prof. Dr. Ulan BRIMKULOV

Üye Doç. Dr. Rayimbek SULTANOV

ONAY

Bu tezin kabulü Enstitü Yönetim Kurulununtarih ve sayılı kararı ile onaylanmıştır.

...../...../.....

Prof. Dr. Zafer GÖNÜLALAN
Enstitü Müdürü

ÖNSÖZ

Çalışmalarım boyunca farklı bakış açıları ve bilimsel katkılarıyla beni aydınlatan, yakın ilgi ve yardımlarını esirgemeyen ve bu günlere gelmemde en büyük katkı sahibi sayın hocam Doç. Dr. Bakıt ŞARŞEMBAYEV'e teşekkürü bir borç bilirim.

Deneysel çalışmalarım sırasında karşılaştığım zorlukları aşmamda yardımlarından ve desteklerinden dolayı bütün hocalarıma teşekkür ederim.

Ayrıca, çalışmalarım süresince sabır göstererek beni daima destekleyen ve bugünlere gelmemde üzerimde büyük emekleri olan aileme sonsuz teşekkürlerimi sunarım.

Meerim UYÇUBEK KIZI

Bişkek, Mayıs 2015

MANAS DESTANINDAKİ METİNLERİN KONUTLARINA GÖRE HYPERTEXTLİK SÖZCÜK DÜZENLEMEK

Meerim UYÇUBEK KIZI

Kırgızistan Türkiye Manas Üniversitesi, Fen Bilimleri Enstitüsü

Yüksek Lisans Tezi, Mayıs 2015

Tez Danışmanı: Doç.Dr. Bakıt ŞARŞEMBAEV

KISA ÖZET

Son yıllarda bilgi teknolojilerinin hızlı gelişimi tüm sektörlerde yararlı oldu ve insan yaşamın kolaylaştırdı. Bununla birlikte her bir ulusun kendi anadillerinin gelişimin ve değişimlerin araştırmada dilsel külliyyatın yardımıyla gerçekleştirmeye başladı. Günümüzde sözlükleri hazırlanma, grameri geliştirme gibi bütün dilbilim araştırmalarında dilsel külliyyat kullanılır. Metinleri kendi anadillerinde hazırlama için de büyük dilsel külliyyat veya veritabanı talep edilir.

Dilsel külliyyat yardımı ile dil ve cümleye ait her türlü bilgiler, statistik veriler alınır.

Her türlü kategorideki kullanıcıların ilgisi için dilsel külliyyat kendi konusu olarak dilbilim araştırmalarda kullanılan büyük cümlelerin dizi kullanımının teorik ve pratik mekanizmasına bakılır diyebiliriz. Bundan dolayı dilbilim kırgız halkının “Manas” destanının araştırmada önemli rol oynar.

ANAHTAR KELİMELERİ:

Dilsel külliyyat, ulusal konut, “Manas” destanı, “Manas” destanının konutu

“МАНАС” ЭПОСУНУН ТЕКСТТИК КОРПУСТАРЫНЫН НЕГИЗИНДЕ ГИПЕРТЕКСТТҮҮ СӨЗДҮКТҮ ТҮЗҮҮ

Мээрим Уйчубек кызы

Кыргыз-Түрк Манас Университети, Табигый илимдер институту

Магистрдик диссертация, Май 2015

Илимий Жетекчи: Доц., Др. Бакыт Шаршембаев

КЫСКАЧА МАЗМУНУ

Маалымат технологияларын өнүгүүсү менен бирге адам баласына көптөгөн тармагында жеңилдетүүлөрдү бере алды. Алардын ичинде ар бир элдин сүйлөгөн эне тилдердин да өнүгүүсүн, өзгөрүүлөрүн изилдөөдө да лингвистикалык корпустардын жардамы менен аткарыла баштады. Азыркы учурда сөздүк түзүү, грамматиканы өнүктүрүү сыяктуу баардык лингвистикалык изилдөөлөрдө лингвистикалык корпус колдонулат. Тексттерди өз тилинде изилдөө үчүн да чоң лингвистикалык корпус же база талап кылынат.

Корпустун жардамы менен тил жана сүйлөмгө тиешелүү ар түрдүү маалыматтар, статистикалык берилиштер алына алат.

Корпустук лингвистика өзүнүн предмети катары ар түрдүү категориядагы колдонуучулардын кызыгы үчүн лингвистикалык изилдөөлөрдө колдонулган чоң сүйлөмдөрдүн массивин колдонуунун теориялык негизин жана практикалык механизмдин карайт десек болот.

Ошондуктан корпустук лингвистика кыргыз элинин “Манас” эпосун изилдөөдө маанилүү роль ойнойт.

АЧКЫЧ СӨЗДӨР:

Корпустук лингвистика, улуттук корпус, “Манас” эпосу, “Манас” эпосунун корпусу.

CREATION OF HYPERTEXT DICTIONARY BODY TEXT OF THE EPIC “MANAS”

Meerim Uychubek kyzy

Kyrgyz-Turkish Manas University, Institute of Natural Sciences

Master Thesis, May 2015

Advisor: Associate professor, Dr. Bakyt Sharshembaev

ABSTRACT

In recent years the rapid development of information technologies has brought many useful facilitations in all sectors of human life and activities. The development of the native language of different nations of the world, which for the studying, development and changing the structure of languages more often started to use a linguistic corpus, takes a special place among them. As for now, to the compilation of dictionaries, in studying various aspects of linguistics, mostly everywhere researchers began using a linguistic corpus of texts. The main condition for a full study of the native languages is the creation or possession of a large linguistic corpuses or databases.

Due to the corpus, we can get different textual and language statistics and materials.

Corpus linguistics as its subject takes a branch of applied linguistics, which engaged in the development of the general principles of construction and usage of linguistic corpuses (text corpuses).

Therefore, corpus linguistics plays a big role in the study of the Kyrgyz epos "Manas".

KEY WORDS:

Corpus linguistics, national corpus, epos “Manas”, corpus of the epos “Manas”.

СОЗДАНИЕ ГИПЕРТЕКСТОВОГО СЛОВАРЯ НА ОСНОВЕ КОРПУСОВ ТЕКСТА ЭПОСА «МАНАС»

Мээрим Уйчубек кызы

Кыргызско-Турецкий университет Манас, Институт
естественных наук

Магистерская диссертация, Май 2015

Научный руководитель: Доц., др. Бакыт Шаршембаев

АННОТАЦИЯ

Бурное развитие информационных технологий в последнее время внес много полезных и всевозможных облегчений во всех отраслях жизни и деятельности человека. Среди них, особое место занимает развитие родного языка разных народов мира, которые для изучения, развития и изменения структуры языка стали чаще использовать лингвистический корпус. На данном этапе в составление словарей, исследовании разных спектров лингвистики повсеместно начали использовать лингвистический корпус текстов. Главным условием для полного изучения национальных языков является создание или владение больших лингвистических корпусов или баз.

Благодаря корпусу получают разные текстовые и языковые статистические данные и материалы.

Корпусная лингвистика в качестве своего предмета берет раздел прикладной лингвистики, занимающийся разработкой общих принципов построения и использованием лингвистических корпусов (корпусов текстов).

Поэтому корпусная лингвистика в исследовании эпоса кыргызского народа “Манас” играет большую роль.

КЛЮЧЕВЫЕ СЛОВА:

Корпусная лингвистика, национальный корпус, эпос “Манас”, корпус эпоса “Манас”.

БАШ СӨЗ

Максаты:Манас эпосунун тексттик корпустарын компьютерге киргизүү менен компьютердик жана математикалык лингвистика тармагында иштеген адистерге ар кандай багытта изилдөөлөрдү жүргүзүүгө жардам бере турчу лингвистикалык корпуста түзүү.

Актуалдуулугу: Азыркы бүткүл дүйнөлүк биригишүү доорунда кээ бир этностордун өз эне тилин жоготуп алуу коркунучу алдында турган коопсуздукту алдын ала көптөгөн мамлекеттер өз эне тилин изилдөө чараларын көрүп турган учур. Ошолордун арасында кыргыз эли да дүйнөлүк мурас болуп калган “Манас” эпосун толук бойдон изилдеп дүйнөлүк электрондук китепканаларга кошулуусу кыргыз элине өз эне тилин сактап калууга чон салым кошот.

Тажрыйбалык аймагы: Лингвисттердин изилдөө методдорун ыңгайлуу түргө алып келүү жана автоматташтыруу тармагы.

Изилдөөнүн негизинде айрым статистикалык маалыматтар боюнча макала даярдалып, К.Карасаев атындагы Бишкек Гуманитардык Университетинин “Кыргыз тили жана адабияты” журналына басууга берилди.

Ишти аткарууда биринчиден идеяны берип багыттап турганы үчүн илимий жетекчим Доц. Б. Шаршембаевге терең ыраазычылыгымды билдирем.

Ошондой эле компьютердик инженерия бөлүмүнүн окутуучуларына, кесиптештериме насааттары жана ой–пикирлери үчүн жана дайыма мени колдоп турган үй-бүлөмө чоң ыраазычылыгымды билдиргим келет.

МАЗМУНУ

İNTİHAL YAPILMADIĞINI BELİRTEN İFADE	iii
ИЛИМИЙ ЭТИКАГА ЫЛАЙЫКТУУЛУК ТУУРАЛУУ БИЛДИРҮҮ	iii
YÖNERGEYE UYGUNLUK.....	iv
KABUL VE ONAY	v
ÖNSÖZ.....	vi
KISA ÖZET	vii
КЫСКАЧА МАЗМУНУ	viii
ABSTRACT	ix
АННОТАЦИЯ	x
БАШ СӨЗ	xii
КЫСКАРТУУЛАР	xv
ТАБЛИЦАЛАР ЖАНА СҮРӨТТӨР ТИЗМЕСИ.....	xvi

БИРИНЧИ БӨЛҮМ

КОРПУСТТУК ЛИНГВИСТИКАГА КИРИШҮҮ

1.1. Корпустук лингвистиканын негизги түшүнүктөрү.....	1
1.2. Корпустук лингвистиканын алдын ала пайда болгон лингвистикадагы багыт: картотекадан корпуска.....	6
1.3. Корпустук лингвистиканын тарыхы.....	10
1.4. Корпустун негизги мүнөздөмөсү	13
1.4.1. Корпустун репрезентативдүүлүгү	13
1.4.2. Корпустарды классификациялоо	16
1.4.3. Параллелдик корпустар	22
1.4.4. Оозеки корпустар	26

ЭКИНЧИ БӨЛҮМ

КОРПУСТАРДЫ ТҮЗҮҮ

2.1. Корпусту түзүүдө алдын ала жасалчу иштер	30
2.1.1. Түзүү процесстерин долбоорлоо	30
2.1.2. Булактарды тандап алуу. Тандап алуу критериялары.....	33

2.1.2. Табигый тилдерди иштеп чыгууда негизги процедуралар: токенизациялоо, лемматизациялоо, стемминг, парсинг.	35
---	----

ҮЧҮНЧҮ БӨЛҮМ **ТЕКСТТЕРДИ ИШТЕП ЧЫГУУ ҮЧҮН КОЛДОНУЛГАН** **ПРОГРАММАЛАР**

3.1. Компьютердик лингвистика	38
3.2. Конкорданс	39
3.3. Конкордансерлердин үлгүлөрү	41

ТӨРТҮНЧҮ БӨЛҮМ **МАНАС ЭПОСУ ТУУРАЛУУ**

4.1. “Манас” эпосу тууралуу кыскача маалымат.....	42
4.2. “Манас” эпосунун көлөмү.	43
4.3. “Манас” эпосун изилдөө жана аны изилдөөчүлөр.....	46
4.5 Манас айтуучулар	47

БЕШИНЧҮ БӨЛҮМ **МАНАС ЭПОСУНУН КОРПУСТАРЫН КОЛДОНУУ**

5.1. Манас эпосунун корпустарында сөз формаларынын орточо кайталанышы.	50
5.2. Манас эпосунун корпустарында сейрек колдонулган сөз формаларынын текстти камтуу көрсөткүчү.....	56
5.3. Тамгалардын айрым статистикасы.	61
5.4. Сөздөрдүн узундугу боюнча маалыматтар.....	67
5.5. Сөздөрдүн кайталанышы боюнча маалыматтар.	81
5.6. Программа үчүн берилиштер базасын түзүү.	94
5.7. Программалык кодду түзүү.	94
5.8. Колдонуучуга маалымат.	96
КОРУТУНДУ.	102
Колдонулган булактар	103
ӨМҮР БАЯН	105

КЫСКАРТУУЛАР

ж.б. – жана башка

ж.б.у.с – жана башка ушул сыяктуу

MS – Microsoft

SQL – Structured Query Language

MSSQL - Microsoft SQL

VB - Visual Basic

DB – Database

PHP – Hypertext Preprocessor (алгачкы аты - Personal Home Page Tools)

Гб – Гигабайт

ЭЭМ – Электрондук эсептөөчү машина

БББС – Берилиштер базасын башкаруу системасы

ББ – Берилиштер базасы

HTTP – Hyper Text Transport Protocol

ТАБЛИЦАЛАР ЖАНА СҮРӨТТӨР ТИЗМЕСИ

Таблица 1.1.1. Корпустук лингвистика менен адаттык (хомскиандык) лингвисткасынын ортосундагы негизги айырмачылыктар.....	5
Таблица 1.4.2.1. Корпустардын классификациялары	17
Таблица 1.4.2.2. Көлөмү боюнча улуттук корпустардын тизмеси	21
Таблица 5.1.1. Сөз формаларынын орточо кайталанышы (<i>толук текст</i>	50
Таблица 5.1.2. Сөз формаларынын орточо кайталанышы (<i>N=10-20-30 миң сөз</i>	51
Таблица 5.1.3. Сөз формаларынын орточо кайталанышы (<i>N=40-70-100 миң сөз</i>	53
Таблица 5.1.4. Эпостордо тексттин көлөмүнө (N) карай сөз формаларынын орточо кайталанышынын (F)өсүшү.....	54
Таблица 5.2.1. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 10 миң сөз.....	56
Таблица 5.2.2. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 20 миң сөз.....	56
Таблица 5.2.3. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 30 миң сөз.....	57
Таблица 5.2.4. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 40 миң сөз.....	57
Таблица 5.2.5. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 70 миң сөз.....	58
Таблица 5.2.6. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (тексттин көлөмү 100 миң сөз.....	58
Таблица 5.2.7. Эпостордо сейрек колдонулган сөз формаларынын квантативдин мүнөздөмөсү (толук текст үчүн.....	59
Таблица 5.3.1. Сөздүк ичиндеги тамгалардын сандык статистикасы (Сөздүн башы)	60

Таблица 5.3.2. Сөздүк ичиндеги тамгалардын сандык статистикасы (Сөздүн ортосу)	61
Таблица 5.3.3. Сөздүк ичиндеги тамгалардын сандык статистикасы (Сөздүн аягы).....	62
Таблица 5.4.1. Сөздүн узундугу боюнча маалыматтар.....	64
Таблица 5.4.2. Баардык вариант үчүн текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	65
Таблица 5.4.3. Манас эпосу Б.Жакиевдин вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	66
Таблица 5.4.4. Манас эпосу Ж.Үмөт уулунун вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	67
Таблица 5.4.5. Манас эпосу Жусуп Мамайдын вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	68
Таблица 5.4.6. Манас эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	69
Таблица 5.4.7. Манас эпосу Б.Юнусалиев 1 вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	70
Таблица 5.4.8. Манас эпосу Б.Юнусалиев 2 вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	71
Таблица 5.4.9. Семетей эпосу Б.Юнусалиевдин вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	72
Таблица 5.4.10. Семетей эпосу Жусуп Мамай 1 вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	73
Таблица 5.4.11. Семетей эпосу Жусуп Мамай 2 вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	74
Таблица 5.4.12. Семетей эпосу Жусуп Мамай 3 вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	75
Таблица 5.4.13. Семетей эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу.....	76

Таблица 5.4.14. Сейтек эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу	77
Таблица 5.5.1. Баардык вариант үчүн текст ичиндеги сөздөрдүн кайталануусу	78
Таблица 5.5.2. Манас эпосу Б.Жакиевдин вариантындагы текст ичиндеги сөздөрдүн кайталануусу	79
Таблица 5.5.3. Манас эпосу Ж.Үмөт уулунун вариантындагы текст ичиндеги сөздөрдүн кайталануусу	80
Таблица 5.5.4. Манас эпосу Жусуп Мамайдын вариантындагы текст ичиндеги сөздөрдүн кайталануусу	81
Таблица 5.5.5. Манас эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн кайталануусу	82
Таблица 5.5.6. Манас эпосу Б.Юнусалиев 1 вариантындагы текст ичиндеги сөздөрдүн кайталануусу	83
Таблица 5.5.7. Манас эпосу Б.Юнусалиев 2 вариантындагы текст ичиндеги сөздөрдүн кайталануусу	84
Таблица 5.5.8. Семетей эпосу Б.Юнусалиевдин вариантындагы текст ичиндеги сөздөрдүн кайталануусу	85
Таблица 5.5.9. Семетей эпосу Жусуп Мамай 1 вариантындагы текст ичиндеги сөздөрдүн кайталануусу	86
Таблица 5.5.10. Семетей эпосу Жусуп Мамай 2 вариантындагы текст ичиндеги сөздөрдүн кайталануусу	87
Таблица 5.5.11. Семетей эпосу Жусуп Мамай 3 вариантындагы текст ичиндеги сөздөрдүн кайталануусу	88
Таблица 5.5.12. Семетей эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн кайталануусу	89
Таблица 5.5.13. Сейтек эпосу С.Каралаевдын вариантындагы текст ичиндеги сөздөрдүн кайталануусу	90

Сүрөт 1 Конкорданс программасынын иштөө чөйрөсү.....	41
Сүрөт 2. Конкорданс программасынын иштөө чөйрөсү(Манас эпосунун варианттарына)	100
Сүрөт 3. Эпостогу сөздөрдү издөө.....	101
Сүрөт 4. Сөздөр боюнча сүйлөмдүн экранга чыгуусу	101
Сүрөт 5. Статистиканын негизги бети.....	102
Сүрөт 6. Сөз узундугу боюнча статистика.....	103
Сүрөт 7. Сөз жыштыгы боюнча статистика.....	104
Сүрөт 8. Саптын саны боюнча статистика.....	105
Сүрөт 9. Орточо жыштык боюнча статистика.....	106

БИРИНЧИ БӨЛҮМ

КОРПУСТТУК ЛИНГВИСТИКАГА КИРИШҮҮ

1.1. Корпустук лингвистиканын негизги түшүнүктөрү

Корпустук лингвистика – бул компьютердик технологияларда лингвистикалык корпустардын (тексттердин корпустарын) жалпы түзүү принциптерин иштеп чыгуу жана колдонуу боюнча компьютердик лингвистиканын бөлүмү.

Лингвистикалык же тексттик корпус катары лингвистикалык маселелерди чечүү үчүн бир түргө келтирилген, түзүлгөн, белгиленген, тексттердин филологиялык жактан толукталган жыйыны болуп саналат.

Азыркы учурда “корпустун” көптөгөн аныктамалары бар. Мисал катары Э.Финегандын китебинде келтирилген аныктамада: корпус – бул машине окуучу форматта, автор, угуучу аудиториялар жөнүндө маалыматтарды камтыган атайын дайындалган тексттердин жыйыны деп берилет.

Википедия баракчасында болсо корпусту тилдин белгиленген жактарда сөздөрдү кезиктирүү же аныктоодо статистикалык изилдөө, текшерүү үчүн электрондук түрдө чоң жана структураланган тексттердин жыйыны деп келтирилет.

Т. МакЭнери жана Э. Вилсон корпус – улуттук тилдердин моделдери катары колдонуу үчүн тилдин так критериялары менен таңдалып алынган тил фрагменттеринин жыйыны деп аныкталат.

Келтирилген аныктамаларда заманбап тексттик корпустардын эң башкы белгилери белгиленген – максат (логикалык идея), машине окула турчу формат, статистикалык жана лингвистикалык маалымат боюнча атайын тандап алуу процедурасы. Сөздүк материалдардын стандартташтырылуусу машиненин жардамы менен иштеп чыгуу үчүн стандарттык жабдыктарды колдонууга мүмкүндүк берет.

Корпустун мааниси, пайдалуулугу төмөнкү шарттар менен аныкталат:

1) Чоң көлөмдөгү (статистикалык түрдө) корпус улуттук тилдердин ар тараптык көрүнүштөрүн берүүгө жана берилиштердин окшош типтүүлүгүнө кепилдик берет;

2) Ар түрдүү типтеги берилиштер өзүнүн контексттик формасында болгондуктан алардын ар тараптуу жана объективдүү изилдөөгө мүмкүнчүлүк берет;

3) Бир жолу даярдалган берилиштердин массиви (корпус) бир нече жолу, ар түрдүү максатта, ар кыл изилдөөчүлөр тарабынан колдонуу мүмкүнчүлүгүн берет.

Тексттик корпус түшүнүгү акыркы убактарда корпустук менеджер (же корпус менеджери) аталган тексттик жана лингвистик башкаруу системасын да камтыйт. Бул атайын багыттагы колдонуучуларга ыңгайлуу түрдө статистикалык маалыматтарды издөө боюнча программдык жабдыктарын камтыган издөө системасы.

Корпуста издөө ар бир сөзгө конкорданс – изделген сөздүн ар бир колдонулган булагына шилтеме түзүүгө мүмкүнчүлүк берет.

Корпус ар түрдүү статистикалык маалыматтарды издөөдө колдонулат. Ошону менен бире корпусун негизинде сөздүн, лексемалардын, грамматикалык категориялардын жыштыгын, алардын белгилүү бир убакыт ичинде колдонуу жыштыгынын өзгөрүүсүн жана башка критерияларды аныктап алса болот. Чоң көлөмдөгү улуттук тилдердин жыйыны аныкталган убакыт ичинде тилдердин лексикалык түзүлүшүн алмашуу процесстерин динамикасын, авторлордун ар кыл жанрдагы лексикограмматикалык изилдөөлөрдү аткарууга мүмкүнчүлүк түзөт.

Корпустар кошумча ар түрдүү тарыхый жана башка сөздүктөрдү даярдоодо көп аспекттүү лексикографикалык иштердин булагы жана аспабы катары да кызмат кылуу функциясын камтыйт. Корпустун маалыматтары улуттук тилдерди окутууда анын грамматикасын толуктап түзүүдө колдонулат.

Азыркы учурда корпусдук лингвистиканы көптөгөн жерлерде компьютердик жана электрондук корпустардын жардамы менен тилди изилдөө боюнча лингвистиканын жаңы агымы деп түшүндүрүлөт. Корпусдук лингвистика өзүнчө дисциплина экендигине эң аз дегенде эки өзгөчөлүгү бар:

- 1) колдонулган сөздүк материалдардын мүнөзү;
- 2) жабдыктардын спецификасы.

Лингвистиканын синтаксис, семантика жана социалдык лингвистика бөлүмдөрү тилдин колдонуусу же тилдин структурасын баалоо же сүрөттөө болсо, корпусдук лингвистика көптөгөн тилдик изилдөө аспектилеринде колдонулган кеңири түшүнүк методология болуп эсептелинет. Корпусдук лингвистиканы кээ бирде “ар тараптуу

лингвистикалык изилдөөнүн методдорунун тутамы” деп да аташат. Лингвистикалык анализдин методу катары корпустук лингвистика салыштыруу изилдөөлөрдө тилдердин, диалектикалардын жалпы жана айрым фактыларын аныктоо катары “контрастивдик” (салыштыруу) изилдөө менен байланыштуу. Лингвистикалык анализдин көптөгөн түрлөрү туруктуу жана эмпирикалык берилиштер базасынын кеңдигине карай жетилүүдө.

Э. Финеган корпустук лингвистиканы тилдин табигый колдонулушун изилдөөгө багытталган корпусту түзүү жана колдонуу боюнча ишкердиги деп аныктайт. Бул аныктамада корпустук лингвистиканын жаратуу багыты белгиленген. Корпустук лингвистиканын оомалык мүнөзү (тексттик корпусту жаратуу менен эле колдонуу максаты) тексттик корпустун оомалык мүнөзү менен шартталат. Бир тараптан башка лингвистикалык дисциплиналар үчүн жана корпустук лингвистиканын баштапкы сөз материалы катары болсо, башка тараптан корпустук лингвистиканын ишмердүүлүгүнүн натыйжасы болуп эсептелинет. Корпустук лингвистика өзүнүн предмети катары көптөгөн колдонуучулардын лингвистикалык изилдөөлөрүнүн теоретикалык негизи, жаратуунун практикалык механизмдери жана берилген улуттук тилдер берилиштер массиви деп айтсак болот.

Кыргыз тилинде корпустук лингвистиканын термини менен байланышкан маселе төмөнкү себептер менен негизделген: АКШ жана Улуу Британияда алгачкы болуп жаралуусу менен терминдердин көпчүлүгү англис тили менен толуктанууда. Кыргыз терминдери негизинен англис жана орус тилдеринен өздөштүрүлгөн.

Корпустук лингвистика анын адаттык лингвистикадан эмне айырмасы бар билишибиз керек. Аны биз төмөндөгү таблицадан көрө алабыз.

Таблица 1.1.1. Корпустук лингвистика менен адаттык (хомскиандык) лингвистикасынын ортосундагы негизги айырмачылыктар

Корпустук лингвистика	Адаттык лингвистика
Негизни багыты – сөздөрдү изилдөө	Негизги багыты – тилди изилдөө
Максаты – атайын түзүлгөн тексттик корпус аркылуу сөздөрдүн колдонуусунун негизинде тилди сүрөттөө.	Максаты – тилди сүрөттөө жана түшүнүү.
Изилдөөлөрдө тексттик корпустун берилиштерине таянат	Изилдөөлөрдө теориядан тилдеги фактыларды аныктоого жана түшүнүүгө багытталат.
Сандык (квантитативдик) методдору колдонулат	Сапатык (квалитативдик) методдору колдонулат
Эмпирикалык методдоруна таянат	Рационалисттик методдоруна таянат
Тексттер физикалык маңыз катары каралат	Тексттер абстракция (илмий ой) катары каралат.
Аныкталган бир тилдин грамматикасын түзөт	Тилдин түрдүүлүгүн изилдейт
Формага негизги маани бөлөт	Формага эле эмес мазмунуна да негизги маани бөлөт.
Тексттерди глобалдык келечектеги планда карайт.	Тексттерди локалдык бүгүнкү күндөгү планда карайт.
Ар түрдүү далилсиз жоболорго (догмаларга) байланбастан тексттин кеңири ой-пикирине маани берет.	Атайын белгиленген маселе туудурган тармакты гана анализдейт.
Тыянак катары тексттерде көрсөтүлгөн сөз аракеттеринде алынган байкоолорго таянат.	Изилдөөлөрдө эмпирикалык жана тилдин материалдарын тандөөдө туюп билүү сезимине таянат.
Алгачкы сөз материалдарын иштеп чыгууда ыктымалдык методдору жана статистикасы көп колдонулат.	Логикалык ой жүгүртүүгө жогору баалайт.
Лингвистикалык берилиштер (сөз колдонумдар) менен кандай түрдө кездешпешсин өзгөртпөстөн колдонот.	Тексттерден атайын тандалган сөз колдонумдарды тандап кодонот.
Эмпирикалык сөз материалдарын иштеп чыгууда индуктивдик методдорун жогору баалайт жана илимий методдун маңызы деп эсептейт.	Эмпирикалык сөз материалдарын иштеп чыгууда дедуктивдик методдорун жогору баалайт.
Эмпирикалык берилиштерин иштеп чыгууда негизделген илимий ачылыштарга ишенет.	Баалоо, салыштыруу жана башка процедураларга негизделген илимий ачылыштарга ишенет.

1.2. Корпустук лингвистиканын алдын ала пайда болгон лингвистикадагы багыт: картотекадан корпуска.

Корпустук лингвистиканы лингвистикадагы эмпирикалык маалыматтарга тиешеси бар методдор, процедуралар жана ресурстар топтому түрүндө элестетүүгө болот. Заманбап корпустук лингвистиканын методология ролунда көтөрүлүшү лингвистиканын тарыхы менен эмпирикалык илимде тыгыз байланышта.

Корпустук лингвистикада колдонулган технологиялар электрондук компьютерлерден алда канча улуу: алардын көбү тамырлары менен лингвистиканын биринчи "реалдуу" же эмпирикалык илим катары жарыялаган XVIII-XIX кылымдарга кетет. Корпустук лингвистиканын негизине жаткан лингвистикалык изилдөөлөрдүн көптөгөн аймактарынын ичинен үчөө гана каралат. Бул үч аймакта технологиянын колдонуусу заманбап корпустук лингвистиканын өнүгүшүнө таасир этет жана тескерисинче.

1.Тарыхый лингвистика:тилдеги өзгөрүүлөр жана реконструкция.

Заманбап корпустук лингвистикага таасир эткен багыттардын бири салыштырма-тарыхый тил илиминен келген. Бул таң калычтуу эмес, себеби тарыхый изилдөөлөр менен алектенген лингвисттер тексттерди же тексттердин топтомдорун негизги күбөлүк катары колдонушкан. Байыркы тилдерди реконструкциялоо же тилдер арасында байланыш куруу учун XIX кылымда өнүккөн көптөгөн технологиялар азыркы күнгө чейин колдонулуп келет. Индоевропалык салттарда тилдеги өзгөрүүлөр жана реконструкциялоо аракеттери биринчи болуп келген тексттерден же корпустардан көз каранды болчу. Я.Гримм жана кечирээк

младограмматиктер тарых жана тилдин грамматикасы жөнүндө ойлорун тексттеги цитаталар менен жактап турушкан.Младограмматиктер өзүнүн манифестинде диалектиде аныкталган заманбап тилди изилдешкенин жарыялашкан жанабул нерсе чоң мааниге ээ болгон.

XIX кылымдан бери өнүктүрүлүп келип жаткан көптөгөн идеялар жана технологиялар корпустук лингвистика тарабынан колдонулган жана өнүктүрүлгөн.Тарыхый корпустарды түзүүмурункудай эле чоң кызыгууларды жараткан.Чынында, электрондук түрдө жеткиликтүү болгон биринчи корпустардын арасында тарыхый корпустарда болгон.

Электрондук форматта пайда болгон тексттердин санынын көбөйүшүкөптөгөн маалыматтарды топтоогожардам берди.Бул нерсе лингвисттерге лингвистикалык анализде статистикалык методдор жардамы менен жеңүүгө жана изилдөөүчүн жаңы методдорду, моделдерди иштеп чыгарууга, өнүктүрүүгө мүмкүнчүлүк берди.Бүгүнкү күндө тилдегиөзгөрүүлөрдү татаал математикалык моделдерин электрондук корпусу аркылуу эсептеп чыгарууга болот.

2.Грамматиканын, лексикографиянын жазылуусу жана тилди үйрөнүү.

XIX кылымдын грамматистери өз ойлорун белгилүү авторлордун чыгармаларынан алынган мисалдар менен иллюстрациялашкан.Мисалы,Г.Пауль өзүнүн немецтик грамматикасында немец классиктеринин чыгармаларын колдонгон.Ал ар өз оюн фонология, морфология жана синтаксис аймактарында иллюстрациялаган.Бүгүн грамматика түзүүчүлөр корпустук подходду колдоно алышат, бирок корпустар классикалык

гана эмес, башка типтеги тексттерди да колдоно алышат.Ошону менен бирге, азыр оозеки сөздүн грамматикасына чоң кызыгуу артууда.Тилдин грамматикалык баяндоосунда корпустардагы түрдүү варианттарды, регистрлерди колдонуунун характеристикасынын жыштыгы жөнүндө маалымат алуу учун колдонууга болот.

Лексикографиядан бир нече алгачкы мисалдарды алсак XVIII-кылымдын ортосунда С.Джонсон англис тилинин түшүндүрмөлүү сөздүгүн жазган.Ал англиялык авторлордун сөздөрдү колдонуу жолун мисалдарда көрсөтүүчүн китептерден иллюстративдүү сүйлөмдөрдү тандап, аларды цитата деп атаган.Окуп жатканда Джонсон сүйлөмдөрдүн астын сызчу жана сөздүн мааниси контексттин ичинде айрыкча түшүнүктүү болуп калчу.Анын ассистенттери асты сызылган сүйлөмдөрдү жаңы кагазга көчүрүп алышчу, Джонсон сөздүктө сөздүк статьяларды түзүү, иллюстрациялоо үчүн ал сүйлөмдөрдү бөлүштүрчү.Сэр Джеймс Муррейдин кол астындагы проект (Англис тилинин Оксфорддук сөздүгү- OED) миңдеген окурмандарды өзүнө тартты жана түзүүчүн жарым кылым талап кылды.

Сенек тилдердин көптөгөн сөздүктөрү контекстинде сөздү камтыган тексттердин цитаталарын берчү.Заманбап корпустук лингвистикада бул метод KWIC(Key Word In Context) корканданс деп аталат.Компьютерлер мисалдарды издөө , классификациялоо , көп сөздүү бирдиктерди бөлүү процесстерин жеңилдеткенине карабастан компьютердик технологиялар жок убагындагы алгачкы лексикографтар жана филологдор колдонгон корпустук тексттердин колдонуу идеялары азыркы убакка чейин абдан окшош.

Адаттагы мектептин грамматика же китептер көбүнчө жасалма түзүлгөн же редакцияланган тилдик колдонуунун мисалдары менен иллюстрацияланган. Келечекте алар студентке жардам бере албайт, анткени студент эртеби кечпи өзүнүн тапшырмаларында, реалдуу баарлашууларында тилдин реалдуу маалыматтарына дуушар болот. Бул учурда корпустар эмпирикалык маалымат булактары катары лингво- дидактикада башкы ролду ойнойт. Тил үйрөнүүдө корпустар студенттин кызыгын ойготуу, аларды аутенттик тилдик колдонууну өз алдынча үйрөнүүгө үндөгөн булак болуп берет. Корпустук маалыматтардын чоң маанилүү колдонуусу - Computer- Assisted Language Learning(CALL) студент тарабынан компьютер аркылуу жасалып жаткан, корпуска негизделген программалык камсыздоо интерактивдүү окуу ишмердигин колдоо учун колдонулат.

3. Социолингвистика: тилдеги көп түрдүүлүк.

Вариативдуу лингвистика XIX кылымда диалектикалык карталарды жана диалектикалык туюнтмалардын жыйнактарын түзүүдөн башталган. Анын методдору ошол кезде колдонулган тарыхый лингвистикалык методдоруна бир айырмалоочу маанилүү белги менен окшош болчу: диалектер корпусу системалык түрдө белгилүү критериялар менен түзүлчү. Балким аны корпуска эмне киргизүү керектиги жөнүндө уланып жаткан дискуссиянын алдын ала билдиргичи катары кароого мүмкүн. Азыркы убакта электрондук корпустар тилдик көп түрдүүлүктүн(мисалы: диалектилердин, социолектердин, регистрлердин) изилдөөлөрүндө жыш колдонулат. Математикалык методдор(мисалы: мультифактордук анализ) толугу менен мындай маалыматтардын жеткиликтүүлүгүнө ишенишет.

Заманбап корпустук лингвистика бул методдорду колдонот жана өнүктүрөт. Көптөгөн изилдөөлөр жана жыйынтыктар электрондук түрдө жеткиликтүү, чоң көлөмдөгү тексттерди колдонуу жана заманбап компьютердик технология аркылуу ишке ашат. Табигый тилдин тексттерин иштетүүчүн жаралган заманбап интеллектуалдык программалык системалардын өнүгүүсү чоң эксперименталдык лингвистикалык базаны талап кылат. Корпустук маалыматтарга суроо-талап техникалык мүмкүнчүлүктөрдүн пайда болуусу менен дал келди.

1.3. Корпустук лингвистиканын тарыхы

Лингвисттер биринчи компьютерлештирилген тексттердин корпусун 1960-чы жылдары чогултушкан. Биринчилерден болуп компьютерлештирилген корпус болуп Броун корпусу (The Brown Corpus) болгон. Ал өзүнө 1961-чи жылы АКШда чыккан америкалык китеп, гезит жана журналдардан алынган 500 текст камтыйт. Броун корпусунда ар бир текст 2000 сөздөн туруп баардык жыйын 1 миллион сөздөн турат (500 текст 2000 сөздөн турат).

Корпустун авторлору У. Френсис (W. Francis) жана Г. Кучера (H. Kucera) жыштыктык, тамга жыштыгы боюнча сөздүк жана башка статистик бөлүүлөр менен баштапкы статистикалык анализдөө материалдары менен коштогон.

Броун корпусун жаратуунун негизги максаты болуп ар кыл англис жазуу жанрларын системдик түрдө изилдөө жана салыштырууларды камсыздоо болчу.

Анын жаралышы жандуу дискуссия менен кызыктурууларды жараткан. Негизинен корпусту чече турган маселелеринин составы жана текстерди тандоо принциптери болгон.

Бир тараптан караганда статистикалык процедураларын негизинде болсо, башка тараптан статистика профессионалдык сезимге таянган автордун эрктүү чечимдеринин негизинде түзүлгөн.

Татаал процесстин максималдуу объективдүүлүгүнө жетүү үчүн максималдуу формалдаштырылган, тунук текшерүү жана контролдоо процедураларды түзүү талап кылынган.

Кийинки европалык изилдөөчүлөр Улуу Британияда 1961-жылы жарык көргөн биринчи тексттик корпусту 2000 сөздөн турган 500 текстти, 15 жанрды ушул принциптин негизинде түзүшкөн. Ал англис сөздөрүнүн британ вариантындагы 1 млн. сөздү камтып Ланкастер-Осло-Берген (The Lancaster-Oslo-Bergen Corpus) корпусу аталган. Браун тибиндеги балансталган корпустар лингвистикалык жактан лингвистикалык баяндоо жана анализдөө үчүн изилдөөчүлөр үчүн абдан манилүү болгон.

Ошентип биринчилерден болгон эки алгачкы корпустар – англис тилинин америкалык жана британ вариантынын жазуу түрүндөгү корпустар болуп эсептелинет. Эки корпус тең пайдалуу болуп калууда. Алардын негизинде көптөгөн англис тилинин изилдөөсү уюштурулган. Корпус жаралгандан бери компьютердик жабдыктардын түрү, күчтүүлүгү артып корпустарды жаратуу процессин оңойлоткон. Азыркы учурда корпустарда миллиарддаган сөздөр камтылган.

1990 жылдары 600дөн ашык компьютердик корпустар катталган. Жасалган жылдары боюнча төмөнкүчө жайгашкан:

-1965	10
1966-1970	20
1971-1975	30
1976-1980	80
1981-1985	160
1986-1990	320

Көрсөтүлгөн статистика боюнча акыркы жылдары жасалган корпустардын саны жана түрү да күчөгөн.

Азыркы заманбап англис тили корпустардын ичинен белгилүү болгон Британ улуттук корпусу (British National Corpus – BNC), англис тилинин эл аралык корпусу (International Corpus of English – ICE), англис тилинин лингвистикалык банкы (Bank of English), заманбап америкалык англис тилинин корпусу (Corpus of Contemporary American English – COCA) жана башкалар. Азыркы учурда дүйнө жүзүндөгү көптөгөн тилдер үчүн корпустар жасалган.

1990-жылдардын биринчи жарымында корпустук лингвистика тил жөнүндөгү билимдин өзүнчө багыты болуп түзүлгөн.

1992-жылы Я. Свартвик алгачкы корпустук лингвистика боюнча Нобелдик симпозиумуна даярдалган маалыматын “Корпустук лингвистика бойго жетилди” деп ат койгон.

Корпустук лингвистика компьютердик лингвистиканын жетишкендиктерин колдонуп ошол эле учурда аны байытып келүүдө.

1.4. Корпустун негизги мүнөздөмөсү

1.4.1. Корпустун репрезентативдүүлүгү

“Корпус” термини акыркы аныкталган көлөмдөгү тексттердин жыйыны деп айтылып келет. Убакыттын өтүшү менен корпустун көлөмү жана түзүлүшү өзгөрүшү мүмкүн, бирок бул өзгөрүүлөр анын түзүлүшүнө залакасын тийгизбеш керек же болбосо түп түкүрү менен өзгөртүүсү зарыл.

Ар түрдүү мүнөздөмөлөрү боюнча корпустун ар кайсы бөлүктөрүнүн өз ара мамилелерин көрсөткүчтүгү реперзентативдүүлүк деп аталган. Алгачкы корпустардын (Браун корпусу, Ланкастер-Осло-Берген корпусу, Упсал орус тилинин корпусу) көлөмү айтылгандай эле 1 млн. сөздү түзгөн. Мындай көлөм улуттук тилдин көп түрдүүлүгүн көрсөтө алган эмес. Азыркы учурда бул маселени чечиш үчүн улуттук тил эң аз дегенде 100 млн. сөздөн туруусу зарыл. Ошондо улуттук корпус тилдин ар өнүгүү убактысы боюнча, ар түрдүү жанр, аймактык жана социалдык түрдүүлүгүн көрсөтө алат. Ошентип азыркы лингвистикалык изилдөөлөр жана сөздүктөрдү түзүү жумуштары кандай болбосун тексттик корпустун көрсөткөчтүгүн (репрезентативдүүлүгүн) колдонууга багытталган.

Корпустун авторлорунун милдети болуп – тилдин изилдөө багытында жасалган корпустун болушунча көп тексттерди чогултуу болуп эсептелинет. Корпус бул улуттук тилдин же тилдин изилдөөчү бөлүкчөсүнүн кичирейтилген модели деп деле айта алабыз. Репрезентативдүүлүк деп тексттик корпуста маселени чечүүчү тараптын бардык сапатын көргөзө ала турган ар түрдүү убакыттагы, жанрдагы, стилдеги проблемаларды керектүү жана толук кандуу

ачып берүү болуп саналат. Репрезентативдүүлүктү аныктоонун ар кандай жолдору бар.

Корпустук лингвистика эң аз дегенде эки типтеги объекттер (тексттик корпустар) менен иш алып барат:

1 Биринчи типтеги корпустар универсалдуу. Алар сөз иш аракеттердин ар түрдүүлүгүн чагылдырат.

2 Экинчи типтеги корпустар коомдук сөз иш аракетиинин лингвистикалык же маданият феномендерин турмушта колдонуусун чагылдырат. Алар *ad hoc* (атайын максаттуу) болуп макал-лакапар корпусу, гезитте колдонулган политикалык метафор корпусу жана башка атайын максаттуу корпустарда колдонулат.

Эки учурда тең репрезентативдүүлүк тексттик корпуста маселени чечүүчү тараптын бардык сапаты көрсөтүлгөнү боюнча статистикалык баа берүү катары гана каралат. Бирок статистикалык баа берүү критериялар дайыма жалгыз жана аныктоочу боло бербейт. Себеби корпус бул каралган маселенин модели катары көрсөтүлөт. Атап айтканда корпустун репрезентативдүүлүгү алынган тыянактардын тактыгын аныктайт. Бул маселени көлөмү боюнча аз болгон тексттик корпуста көп тексттердин массивин толук көрсөтүү же интеграциялоо маселеси катары да караса болот.

Улуттук тилдер чындыгында ар түрдүү жана ар кайсы фактурада (оозеки, жазуу ж.б.) көргөзүлгөн жана аларда ар түрлүү жазылган лингвистикалык көрүнүштөр абдан көп. Алтымышынчы жылдары биринчи типтеги тексттик корпустар көрсөтүлгөн тилди же өзүнчө бир фрагментин статистикалык жактан туура көрсөтүп универсалдуулукка талаптанган. Мисалы катары Браун корпусу

АКШда алтымышынчы жылдары тилдин басма басуу формасынын жетишээрлик репрезентативдүүлүгү менен жаралган. Алынган тексттер айтылгандай эле 15 жанрды камтыш керек болгон. Алардын ичинен алтыдан сексенге чейинки жанрдын ар кайсыл жеринен тандалып алынган.

- 1) пресса: репортаж;
- 2) пресса: баш макала;
- 3) пресса: обзорлор;
- 4) диний тексттер;
- 5) ыкмалар, кесип, хобби;
- 6) илимий-популярдуу адабият;
- 7) беллетристика, биографии, эссе;
- 8) ар кайсы(өкмөттүк документтер, ишканалардын отчеттору ж.б.);
- 9) илимий сочинениялар;
- 10) көркөм адабият;
- 11) мистика жана детективдер;
- 12) илимий проза;
- 13) укмуштуу окуялуу адабият жана вестерндер;
- 14) сүйүү романдар;
- 15) юмористик чыгармалар.

Экинчи типтеги корпуста болсо репрезентативдүүлүк критерий катары кубулуштарды максималдуу объективдүү көрсөтүүсү болуп саналат. Мисалы англис тилинин бир убакытта жана географиялык аймакта колдонулган максималдуу репрезентативдүү англис макал-лакаптар корпусу англис политикалык метафорлорун изилдөөгө репрезентативдүү боло албайт.

XX кылымдын башында гезиттин башкы аты корпусу, англис тилинин корпусу, политикалык метафор корпусу сыяктуу корпустар талкууланган. Корпустун негизинде ар дайым карала турчу илимий изилдөөлөрдүн маселесин уюштуруу болуп саналган. Ошондуктан корпуска тексттерди тандап алууну түзүүчү өзүнүн практикалык же илимий ишкердүүлүгүнүн максатына жараша белгилеген.

Корпусту түзүү методдору корпустун тибине байланыштуу болот. Бул маселе актуалдуу жана жетишээрлик изилденген эмес. Биринчи типтеги корпусту түзүү методологиясы жалпы маселени чагылдырган жеке бир тексттик корпуста көргөзүлүп дедукция принцибине негизделет. Экинчи типтеги корпустун методологиясы лингвистикалык феномендердин айрым учурларын атайын ошол учурларды көргөзө турган корпуска негизделет. Бирок теория менен практика көрсөткөндөй эки ыкма бирдикте колдонулуп келет.

1.4.2. Корпустарды классификациялоо

Корпус түрлөрүнүн көптүгүнө карабастан алардын эки класска бөлүү ыкмасын бөлүп айтса болот:

1) Улуттук тилдин (көбүнчө белгиленген убакыттагы тилдин) корпусун ошол эле тилге тиешелүү бөлүмдөрүнүн (жанр, стиль, ж.б.) корпустары менен салыштыруу ыкмасы;

2) Корпустарды лингвистикалык белгилөө тиби боюнча. Көп түрдөгү белгилөө типтерин болуусуна карабастан корпустардын көбүнчөсү морфологиялык жана синтаксистикалык типке киришет. Ошондой эле айта кетчү нерсе болуп синтаксистикалык белгилөөсү

бар корпус ачык же түнт сөздүн морфологиялык мүнөздөмөлөрүн да камтып кетет.

Жалпысынан алганда көптөгөн изилдөөлөргө ылайыкталып ар кайсы классификациялардын негизинде жасалган көп сандагы ар түрдүү корпус бар. Астыга коюлган максатка жана классификациялык белгилерге жараша ар типтеги корпустарды бөлүп чыкса болот. (табл.1.4.2.1.)

Таблица 1.4.2.1.Корпустардын классификациялары

Белги	Корпустардын типтери
Тилдик берилиштер тиби	Жазма Оозеки Аралашкан
«Параллелдик»	Бир тилдүү Эки тилдүү Көп тилдүү
«Адабияттык»	Адабияттык Диалектик Оозекилик Терминдик Аралашма
Максат	Көп максаттуу Атайын багыттуу
Жанр	Адабий Фольклор Драма чыгармаларыкоомдук- саясий боюнча
Жеңилдик	Анчалык көтөрүнкө эмес коммерциялык жабык
Багыты	Изилдөө Көрсөтмөлүү
Динамика боюнча	Динамикалуу Статистикалуу

Белги	Белги урулган Белги урулбаган
Белгинин харакетри боюнча	Морфологиялык Синтаксикалык Семантикалык Просодикалык ж.б.
Тексттин көлөмү боюнча	Толук тексттүү «Фрагменттүү тексттер»

Ошентип, *тилдик берилиштер тиби* боюнча корпустар жазма, оозеки жана аралашкан болуп бөлүнүшөт. Жазма корпустарда оозеки корпустар көргөзүлгөн эмес (Браун корпусу, LOB), оозеки корпустарда оозеки корпустар эле берилет. Аралашкан корпустар болсо бир убакыт ичинде чогултулган улуттук корпустар болот (НКРЯ, BNC ж.б.).

Параллелдик критериясы боюнча корпустар бир тилдүү, эки тилдүү жана көп тилдүү болуп бөлүнүшөт. Бир тилдүү корпустарды тилдердин варианттары, диалекттер салыштырылат. Мисалы катары англис тилинин эне тили катары жана башка тилдерден келген чет өлкөлүк англис тил катары контрастивдик анализ үчүн салыштыруу процесси. Эки тилдүү жана көп тилдүү корпустар эки же көп тилде жазылган окшош тематикалык жазуулар болуп эсептелинет. Мисал катары бир нече өлкөлөрдө ар кайсы тилде өткөн илимий конференциянын материалдардын корпусу. Мындай корпустар терминдер менен иштеген котормочулар тараптан көп колдонулат. Дагы бир мисал катары бир тилде жазылган чыгарманын башка көптөгөн тилдерге которулган варианттары жанаалардын салыштырылмасы боло алат. Мындай корпустар салыштырмалуу

изилдөөлөрдү жасоого, которуу теорияларын изилдөөдө баа жеткис материал катары саналат.

Адабияттык критериясы боюнча адабияттык, диалектик, оозеки, терминдик жана аралашма корпустарына бөлүнөт. Оозеки корпусу катары Санкт-Петербург шаарында түзүлүп жаткан Один Речевой День (ОРД) корпусу, терминдик корпус катары терминдик сөздүк түзүүдөгү корпустук лингвистика боюнча тексттик корпус саналат.

Корпусту түзүүдө *максаты* боюнча көп максаттуу жана атайын багыттуу болуп бөлүнөт. Көп максаттуу корпустар дайыма ар түрдүү жанрдагы тексттерди камтыйт. Атайын багыттуу корпус болсо бир же бир нече жанр менен гана чектелет.

Тексттик корпустар жанр боюнча да классификацияланып адабий, фольклор, драма чыгармалары, коомдук-саясий корпустарга бөлүнүшөт. Коомдук-саясий корпус катары XX-чы кылымдын орус гезитинин компютердик текст жана саясий метафор корпустары боло алат.

Корпустун колдонуучуларынын маанилүү критерийи катары анын *жеңилдиги* болуп саналат. Эркин жетишээрлик корпустар баардык тексттик корпуска он-лайн режиминде толук көлөмдө каалаган убакытта колдонууга мүмкүндүк берет. Коммерциялык корпустарды колдонууда он-лайн же компакт дисктин колдонуу үчүн колдонуу укугун сатып алуусу керек. Жабык корпустар болсо тар чөйрөдө колдонуу үчүн түзүлүп эл арасында колдонууга максатталган эмес.

Багыты боюнча изилдөө жана көрсөтмөлүү корпустарга бөлүнүшөт. Изилдөө корпустары тилдин иштөө аспектерин изилдөө үчүн түзүлөт. Бул типтеги корпустар лингвистиканын кеңири маселелерин изилдөөгө багытталган. Көрсөтмөлүү корпустар илимий изилдөөлөрдү аткаргандан кийин түзүлөт, себеби алардын максаты жаңы фактыларды издөө эмес, алынган жыйынтыктарды түшүндүрүп негиздөө болуп эсептелинет. Мисал катары бөлүктөрүн семантикалык анализин жана көрсөтүлгөн мисалдарды текшерүү үчүн чоң көлөмдөгү материалдар менен коштолгон «Путеводителе по дискурсивным словам русского языка» корпусу саналат.

Динамика боюнча критериясы динамикалуу жана статистикалуу болуп экиге бөлүнөт. Башында корпустар статистикалуу болуп улуттук тилдин убактылуу абалын гана көргөзгөн. Статистикалык корпус белгиленген убакыт арасындагы тексттерди гана камтыйт. Мисал катары жазуучулардын автордук корпусу белгиленет. Бирок көпчүлүк лингвистикалык маселелердин максаты болуп белгиленген убакыт ичиндеги тилдин өзгөрүү кубулуштарын чагылдырып алардын колдонуу жаштыгын, синтаксистикалык өзгөрүүлөрүн аныктоо болот. Мындай маселелерди чечүү үчүн жаңы динамикалык корпустар түзүлүп чыккан. Динамикалык корпустар монитордук же болбосо байкоолуу деп аталган. Монитордук корпустардын максаты жылдан жылга көбөйгөн тексттерди компьютердин эсине чогултуу болот. Убакыттын өтүшү менен тексттер жаңыланып көптөгөн тексттер кошулуп турат. Дайыма өнүккөн мониторинг корпустар стилдерге жараша сөздөрдү колдонуу балансын, жаңы кирген сөздөрдү же болбосо маанисин өзгөртүүсү боюнча лексикографтардын жумушун

онойлотуп, ар түрдүү диахроникалык изилдөөлөрдү аткарганга жана жаңы сөздүктөрдү түзүүдө абдан чоң ролду ойнойт.

Белги критериясы боюнча белги урулган жана белги урулбаган болуп экиге бөлүнүшөт. Корпустун бул критериясы боюнча башка да индекстелген жана индекстелбеген, аннотациялаган жана аннотацияланбаган, тектүү жана тексиз деген да аталыштары кездешет. Белгиленген корпуста сөздөргө же сүйлөмдөргө белгинин характерине жараша морфологиялык, синтаксикалык, семантикалык, просодикалык сыяктуу белги (тэг) ыйгарылат.

Тексттин көлөмү боюнча критериясы толук тексттүү жана фрагменттүү тексттерге бөлүнөт. Белгилүү болгондой Браун жана Ланкастер-Осло-Берген корпустарында тексттин жыштыгы 2000 сөздөн туруусу зарыл эле. Бирок мындай шарттарга так туура келген тексттер аз болгондуктан бул корпустар фрагменттүү тексттерге кирет. Толук тексттүү корпустарга бир автордун корпусу, кыска тексттүү корпустар же болбосо гезит баш тамгалар корпустары каралган.

Таблица 1.4.2.2. Көлөмү боюнча улуттук корпустардын тизмеси

Орус тилинин улуттук корпусу http://ruscorpora.ru	300 млн. Сөз
Британ тилинин улуттук корпусу http://www.natcorp.ox.ac.uk/ или http://sara.natcorp.ox.ac.uk/	100 млн. Сөз
Венгер тилинин улуттук корпусу http://corpus.nytud.hu/mnsz/	100 млн. Сөз
Испан тилинин улуттук корпусу (тарыхый) http://www.corpusdelespanol.org/	100 млн. Сөз
Заманбап дат тилинин улуттук корпусу http://www.korpus2000.dk/	50 млн. Сөз
Заманбап итальян тилинин улуттук корпусу http://www.cilta.unibo.it/ricerca.htm	100 млн. Сөз
Заманбап кытай тилинин улуттук корпусу	720 млн. Сөз

(LIVAC Synchronous Corpus) http://www.rcl.cityu.edu.hk/livac/	
Немец тилинин мангеймдик корпусу (Institut für Deutsche Sprache, Mannheim, Germany) http://corpora.ids-mannheim.de/~cosmas/	1610 млн. Сөз (150 млн. Иероглиф)
Словен тилинин улуттук корпусу http://www.fida.net/eng/	100 млн. Сөз
Польша улуттук корпусу http://korpus.ia.uni.lodz.pl	93 млн. сөз
Хорват улуттук корпусу	53 млн. сөз
Чех улуттук корпусу	100 млн. сөз

1.4.3. Параллелдик корпустар

Параллелдик корпустар эки негизги типке бөлүнөт:

1. Бир тилде жазылган оригинал тексттер жана оригинал тексттердин бир же бир нече башка тилдерге которулган которулма тексттердин корпусу;
2. Бир же бир нече тилдерде жазылган тематикалык жактан окшош тексттердин корпусу.

Биринчи жана экинчи типтеги корпустар лексика, грамматика, стиль, котормо жактан салыштырмалуу изилдөөлөрдө колдонулат.

Биринчи типтеги параллелдик корпустарды түзүүдө жана аларды иштеп чыгуучу жабдыктарды түзүүдө оригинал текст менен анын котормосунун шайкештигин далилдөө маселери ортого чыгат. Бул маселени чечүү үчүн тексттерди автоматтык тегиздөө (alignment) методу колдонулат. Бул методдун негизи болуп оригинал жана анын котормосу ортосунда сөздүн жанан сүйлөмдүн грамматикалык түзүмүнө ылайык параллелдүү сегментация эсептелинет.

А.В. Зубов жана И.И. Зубованын китебинде сүйлөмдү тегиздөөдө эки тексттин окшоштугу үчүн алты шайкештикти колдонгон:

- 1) бир оригинал сүйлөм бир сүйлөм менен которулат;
- 2) эки оригинал сүйлөм бир сүйлөм менен которулат;
- 3) бир оригинал сүйлөм эки сүйлөм менен которулат;
- 4) эки оригинал сүйлөм эки сүйлөм менен которулат;
- 5) оригинал текст которулбайт;
- 6) оригинал текст которула турган тилде окшош мааниси жок.

Бул маселелерди чечүү үчүн ортомчу тил же универсалдуу тил колдонуу теориясына негиздесе болот.

Практикада сүйлөмдөрдөн узундугун салыштыруу, абзацтарга бөлүү, тыныш белгирин анализдөө, сөздүктөрдү колдонуу жана башка факторлордун негизинде салыштырып иштеген ар түрдүү тегиздөөчү программдык жабдыктар бар. Бул программалар көбүнчө автоматтык тегиздөөдөн кийинки адам-машине которуу варианттарында колдонулат.

Параллелдик корпустар чоң көлөмдөгү маалымат алууга мүмкүнчүлүк түзөт. Алардын жардамы менен:

- ✓ Эки тилдүү жана көп тилдүү котормо сөздүктөрдү;
- ✓ Машине котормо системасы үчүн сөздүктөрдү түзүү жана топтоо;

- ✓ Компьютердин жардамы менен окшош сөздөрдүн узундугу боюнча лексикалык бирдиктердин полисемиясын чечүү;

- ✓ Тексттин терминдик жана фразеологиялык бирдиктерин которуу;

- ✓ Которуу эси аталган жаңы машине которуу системасынын компьютердин эсинде топтолгон оригинал жана анын которулма тексттер корпусун ар түрдүү деңгээлде текшиленген толук автоматташтырылган котормолорду жазаса болот.

Которуу процессинде мындай система которула турчу сүйлөмдү же фрагментти оригинал параллелдик тексттин массивинде табууга аракет кылат. Эгерде ал сүйлөм оригинал параллелдик тексттин массивинде табылса система сүйлөмдүн котормосун же фрагменттин которулган текстинин массивинен тандайт.

Параллелдик корпустарды изилдөөдө лексиканы автоматтык классификациялоо инструменттерин колдонсо болот. Лексиканы автоматтык классификациялоосу тексттердин анализдөөдө маанилүү процедураса болуп эсептелинет. Ал тексттин түзүлүшүн формалдаштыруу жана тексттин элементтеринин (лемма жана сөз формалар түрүндө) арасында семантикалык байланыштарды сандык баалоо чегинде аткарылат. Сүйлөмдөрдүн колдонуу жыштыгын, семантикалык жакындыгын аныктаган салыштырмалуу анализ лексика-семантикалык жана тематикалык топтун ичиндеги ар түрдүү тилдердин лексикалык бирдиктеринин жайылышын аныктоого жардам берет. Экинчи типтеги параллелдик корпуста оригинал жана котормо тексттердин параллель иштеп чыгуунун

негизиндеги класстер элементтери жөнүндөгү маалымат контрастивдик изилдөөлөрдө жана котормонун окшоштугун аныктоодо чоң роль ойнойт. Лексиканын автоматтык классификация модулдарын колдонуу параллелдик корпустарда издөөнүн эффективдүүлүгүн арттырат, көп тилдүү сөздүктөр үчүн толуктоо жана ондоого берилиштерди байма бай бере алат жана машине аркылуу котормолорду, аларды үйрөтүү үчүн иш сапатын текшергенге жол ачат.

Автоматтык которуу системасы окшош грамматикалык түгөйлөрүн тандоо эрежесин колдонуу менен эки тилдин морфологиялык байланыштарына негизделсе болот. Мисалы орус жана словак тилдерин лексикалык системасынын генетикалык окшоштугу бир тилде жок болгон сөздөрдү транслитерация схемасынын жардамы менен которуу сапатын жогорулатууга салым кошот.

Мындай метод ар түрдүү которуу процесстерин автоматташтырууга мүмкүнчүлүк берет. Машине жардамы менен которуу системасы эки же андан көп окшош тилдердин, мисалы словак жана чештилдердин синтаксикалык окшоштугун колдонуп келет. Параллелдик корпустар БУУ расмий колдонуу тилдери, Бириккен Европа жана расмий түрдө эки тилдүү өлкөлөрдүн тексттердин негизинде түзүлөт.

1.4.4. Оозеки корпустар

Оозеки чыгарманын репрезентативдүү корпусун түзүү оор маселе болгондуктан компьютердик лингвистика жана корпустук изилдөөлөр лингвистиканын башка бөлүмдөрүнө караганда абдан жакшы изилденген.

Лондон-Лунд корпусу (The London-Lund Corpus) «Англис тилин колдонуу» (The Survey of English Usage) долбоорунун чегинде түзүлгөн. Долбоордун максаты болуп англис тилинин грамматикалык системасынын өзгөчөлүктөрүн колдонуу болгон. Долбоор 1960-жылдары Р. Квирктин көзөмөлдөөсүнүн алдында Лондон университетинде түзүлгөн. Корпустун көлөмү 1 млн. сүйлөмдү түзгөн. Оозеки тексттер катары радиоберүүлөрдүн, расмий отуруштардын тексттери киргизилген. Корпустун машине түрү 1979-жылы Швецияда Лунд университетинде жазылган. Лондон-Лунддун оозеки корпусу биринчи машине окуй ала турган корпус болуп түзүлгөн. Ал 34 сүйлөмдөн турган жашыруун жазылган маектин текстти болуп Дж. Свартвик жана Р. Квирктин “Англис тилинин оозеки корпусу” аттуу китебинде да жарыяланган. Бул китеп абдан пайдалуу болгон, себеби ал учурда компьютердик корпустардын жок болгондугунан оозеки тилдин транскрипциясы менен иштеш оор болгон. Оозеки корпусту түзүү оор болгондуктан, алгачкы корпус англис тилинин компьютердик изилдөөлөрү үчүн абдан жакшы булак болуп келген.

Лондон-Лунд корпусунун жарык көрүшү лексика, грамматика, өзгөчө дискурстун түзүлүшү жана функциялары боюнча көптөгөн изилдөөлөргө жол ачкан. Мисалы actually, really, you know, you see, I mean, well сөздөрдүн колодонулушу, англис тилинде суроо-жооп жана башка темалар изилденип чыккан. Англис тилинин оозеки жана жазуу түрлөрү Лондон-Лунд жана Ланкастер-Осло-Берген корпустарынын негизинде салыштырмалуу изилдөөлөр өткөрүлгөн. Азыркы учурда корпустук лингвисттердин чоң кызыгуусун жаратып келе жаткан оозеки чыгармачылыкта эмоцияны, таң калуу жана башкаларды түшүрүү болуп саналат.

Оозеки жана жазуу корпустарынын арасында машине окула турчу формада тең салмактуулук жок. Ар түрдүү себептер менен оозеки корпусту түзүү, жазуу корпусун түзүүгө караганда жай жүрүүдө. Биринчиден оозеки чыгарманы магниттик лента, магнитофон же видеокасета жардамы менен жазыш керек. Андан соң аларды тамга түрүнө түшүрүү оор жумушка айланат.

Лингвистиканын фонетикалык каражаттарын түзүүнүн негизги маселелери оозеки чыгармалардын транскрипцияга түшүрүү болуп эсептелинет.

Бул учурда төмөнкү маселелер ортого чыгат:

1. Транскрипция үчүн кайсы алгоритмди колдонуу керек?
2. Индивидуалдык так айтууну эске алуу керекпи?
3. Оозеки чыгарманын баардыгын же фрагментин эске алуу керекпи?

4. Сөздөрдүн диалектик айырмачылыктарын эске алуу керекпи?

5. Сөздөрдөгү басымга көңүл буруу керекпи?

6. Айтылып жаткан сүйлөмдөрдүн просодикалык мүнөздөрүн эске алуу керекпи?

7. Угулбай калган сөздөрдү эске алуу керекпи?

8. Фонетикалык корпуста паралингвистикалык кубулуштарга (тынуу, күлүү, кобууроо, жөтөлүү ж.б.) көңүл буруу керекпи?

Азыркы учурда машине окулуучу фонетикалык корпустарды түзүүдө үндөрдүн жардамчы белгилер менен жада калса просодикалык, паралингвистикалык мүнөздөмөлөрү менен түзүлгөн транскрипциясы колдонулат. Мындай оорчулукка карабастан дүйнө жүзүндө көптөгөн фонетикалык оозеки корпустар бар. Мисалы XX кылымдын 70-жылдары АКШда Х. Далем жана анын кесиптештери менен түзүлгөн “Англис тилинин америкалык вариантынын оозеки корпусу” бар. Ал корпуста психоаналитикалык сеанстарда чогултулган 1 млн. сөз колдонумдарын камтыйт. Тандалып алынган жазуулар англис тилинин орфографиясын стандарты менен транскрипцияланган. Диалектик айырмачылыктар эске алынган эмес. Угулбай калган сөздөрдү Z тамгасы менен белгилешкен. Сүйлөмдөрдүн просодикалык мүнөздөрү жана басымдары да эске алынган эмес. Бирок ошол эле учурда атайын комментарий катары паралингвистикалык кубулуштар (тынуу, күлүү, кобууроо, жөтөлүү ж.б.) эске алынган.

Британ улуттук корпусун жаратуучулардын бири болгон Л. Бернард оозеки булактардан 10 млн сөздү тандап алуу жазуу булактардан 50 млн сөз тандап алгандай деп баа берген.

Орус тилинин улуттук корпусунда 2008-жылы оозеки тексттер 3,9% гана түзгөн. Анын ичинен оозеки тексттердеги элдик сүйлөм 64,3%, элдик эмес сүйлөм 8,1% жана кинодон алынган сүйлөмдөр 27.6% ды түзгөн.

Ошондуктан улуттук корпустарды түзүүдө чоң маселе катары оозеки чыгармалардын аз болгону, айрыкча элдик эмес сүйлөмдөр (телефон сүйлөшүүлөр, расмий эмес маектер ж.б.) белгиленет.

ЭКИНЧИ БӨЛҮМ

КОРПУСТАРДЫ ТҮЗҮҮ

2.1. Корпусту түзүүдө алдын ала жасалчу иштер

2.1.1. Түзүү процесстерин долбоорлоо

Корпусту түзүүдө анын этаптары жана анын кийинки өнүгүү жолдору алдын ала каралышы керек. Корпус түшүнүгү лингвисттердин картотека менен иштөө процессинин уландысы болуп эсептелинет. XX - кылымда корпустар компьютерлешип эл ичинде тарай баштаган. Корпустун өнүгүүсүнө интернет тармагы да чоң салым кошту. Ошону менен бирге чогултулган материалдардын баланстуулук жана репрезентативдүүлүк суроолору да актуалдуу бойдон калган.

Жанр жана тематикалык түзүлүштөрдөн башка да көптөгөн маанилүү суроолорду чечүү зарыл. Алар төмөнкүлөр:

1. Корпус ичиндеги тексттердин мааниси кандай? Мисалы гезиттердин кыска макала аттары корпуста өзүнчө бирдик болобу же баардыгы бир корпус болобу?

2. Гезитте макала текст катары каралабы? Же болбосо гезиттин өзү бир текст катары каралышы керекпи?

3. Ыр жыйнагы текст катары каралабы же алардын бир куплети текст катары каралабы?

Хронология маселеси да чоң мааниге ээ. Азыркы кыргыз тилинин корпусу эмне түшүнүк бере алат? Ар түрдүү жанрлардын

корпуста хронологиялык чеги болушу керектигин да эске алуусу зарыл.

Корпус ар түрдүү колдонуучулар жана ар түрдүү маселелер үчүн түзүлөт. Корпусту түзүү процессинде баштапкы тексттер тазаланат. Мисалы сүрөт материалдары тил материалдарына кирбегени үчүн тазаланышы керек. таблицадар, цитаталар, башка тилдерден келген сөздөр менен иштөө оор маселелердин катарына кирет.

Ушул сыяктуу суроолор долбоорлор бөлүгүндө каралышы зарыл. Суроолордун кээ бирлери корпусту түзүп жатканда же колдонуп жатканда да акырындык менен чечилиши мүмкүн.

Корпусту түзүү процесстерин төмөнкү кадамдардан турат:

1. Булактардан тексттерди чогултуу.
2. Машине окулуучу формага келтирүү. Тексттерди электрондук түргө ар түрдүү жол менен келтирсе болот. Мисалы: кол менен киргизүү, сканерлөө ж.б. жолдор.
3. Текстти алдын ала анализдөө жана иштеп чыгуу. Бул кадамда ар кайсы булактардан тандалып алынган тексттер филологиялык текшерүдөн өтүп оңдолот.
4. Графематик анализ жана которуу. Кээ бир тексттер машина окулуучу формага келтирүүдө ар түрдүү коддорго которуу, текст болбогон элементтерди (сүрөт, таблицадарды) алып салуу, бир типте тиркемелерди коюу ж.б. процесстер аткарылат. Графематик анализ катары баштапкы тексттерди элементтерге бөлүү (сөздөрдү, бөлүү белгилерди), текст болбогон элементтерди алып салуу, башка тилден

келген лексемаларды, колдонулган булактардын аттарын алып салуу болуп эсептелинет. Бул кадамда тексттер сегменттерге бөлүштүрүлөт.

5. Текстти белгилөө. Белгилөө бул тексттерге жардамчы маалыматтарды(мета берилиштер) кошуу болуп эсептелет. Мета берилиштерди үч типке бөлсө болот: баардык тексттерге тиешелүү экстралингвистикалык; тексттин түзүлүшү жөнүндө маалымат; тексттин элементтерин сүрөттөгөн лингвистикалык мета берилиштер. Тексттик корпустун мета сүрөттөөсү берилиштер элементтердин библиографиялык берилиштер, жанр жана стилдик өзгөчөлүктөрү, автор жөнүндө маалыматтырды, ошондой эле файлдын аттары, коддоо усулу, ишти жасаган адамдардын аттары сыяктуу формалдуу маалыматтарды камтыйт. Бул иштердин баары кол менен кирзилет. Документтен түзүлүш белгилери (сүйлөм, сөз, абзацтарды чыгаруу) жана лингвистикалык белгилөө автоматтык түрдө аткарылат.

6. Автоматтык белгилөө процессинин натыйжасын түзөө иштери: каталарды оңдоо.

7. Белгиленген тексттерди көп аспекттүү издөө үчүн атайын даярдалган лингвистикалык маалыматтык издөө системасына (corpus manager), которуу жана статистикаларын иштеп чыгуу.

8. Корпусту колдонуу мүмкүнчүлүгүн берүү. Корпус компакт дискте же болбосо интернет тармагы аркылуу дүйнө-жүзүнө таралышы мүмкүн. Ар кыл категориядагы колдонуучуларга ар түрдүү колдонуу укугу берилиши керек.

9. Корпусту түзүүдө колдонулган каражаттар, колдонуу усулдары, белгилөө системаларын ачыктап жазылган документтерди даярдоо.

Ар кыл корпусту түзүү процесси ар кандай болгондуктан каралган кадамдардан айырмаланышы мүмкүн.

2.1.2. Булактарды тандап алуу. Тандап алуу критериялары.

Тексттик корпустун маанилүү өзгөчөлүгү катары ар кандай тилдеги ар түрдүү тексттердин жөн эле жыйыны эместиги. Корпусту түзүүдө ар кандай суроолор жаралат.

Ал суроолор төмөндө көрсөтүлгөн:

1. Тексттик корпуста негизги чоңдук эмне болушу керек?
2. Тексттик корпустун көлөмү канча болушу керек (канча чоңдуктан турушу керек) ?
3. Корпустук текстте кайсыл жазуу булактары жана канча көлөмдө болот?
4. Корпустук текстке кайсыл лингвистика тармагындагы тексттер кириши керек?

Бул суроолорго биринчилерден болуп Р.Г. Пиотровский жана анын шакирттеринин 1965-жылдары өткөрүлгөн лингвостатистикалык изилдөө жана сөз жыштыгын жасоодо тексттерди тандап алуу изилдөөлөрүндө жооп берген. Ушул эле суроолор Л.Н. Засорина жетекчилиги алдында жыштык сөздүгүн

түзүүдө талкууланган. Ошол убакта биринчи жолу ар түрдүү тандап алуу көлөмү, үлүшүн аныктоо үчүн статистикалык ыкмалар колдонулган.

Корпустун негизги чоңдугу катары сөз формалары (же болбосо сөздөр), негизи (уңгу, лемма) жана сүйлөм болушу мүмкүн. Түзүлүп жаткан корпустун көлөмү колдонула турчу иштин максатына жараша болот. Тамгаларды колдонуу жыштыгы, үндөрдү изилдөөдө көлөмү аз болуп, морфологиялык кубулуштарды, лексиканы, тексттердин морфологиялык, стилистикалык өзгөчөлүктөрүн изилдөөдө көлөмү чоңоюшу мүмкүн.

Маселе катары төмөнкү суроолор пайда болот:

1. Тексттик корпуста кайсы функционалдык жанрлар тексттерин кошулушу керек (драма, ыр, илимий текст, гезит, журнал ж.б.)?
2. Кайсы убакыт аралыгындагы тексттер кошулушу керек (заманбап, 1- жыл мурунку, 5- жыл мурунку, байыркы тексттер)?
3. Корпуска адабий тексттерди гана же болбосо башка типтеги булактарды кошуу керекпи?

Ушул сыяктуу суроолорго жооп берээрде лингвистика статистикасы боюнча адистерин кеңешин жана анкета методдорун колдонушат. Изилдөө тажрыйбаларына жараша адистер корпустун жалпы көлөмүн, тексттин жаралыш убактысын, тандоо тексттеринин санын, көлөмүн, жанрын аныкташат. «Американский корпус наследия» (The American Heritage Intermediate Corpus) түзүүдө анкета жана адистердин кеңеши колдонулган. Корпус 5 млн сөздө турганын аныкташып кошумча англис тилинде балдар жана

өспүрүмдөр адабиятынын 22 жанрын кошууну сунушташкан. АКШда 221 мектепке корпуста кайсыл тексттер болушун каалаганын билиш үчүн анкеталар таратылган. Анкетаны негизинде 19 миң китептин тизмеси түзүлгөн. Булардын арасынан 1045 текст тандалып алынган. Натыйжада 500 сөздөн турган 10 миң жөнөкөй сурактар (выборка) даярдалган.

2.1.2. Табигый тилдерди иштеп чыгууда негизги процедуралар: токенизациялоо, лемматизациялоо, стемминг, парсинг.

Корпус бул компьютердин берилиштер базасы жана аны түзүүдө атайын процедуралар, программалар жана анын мыйзам түрүндө колдонулушу. Мисалы: **токенизациялоо**, б.а. символдордун агымын маанилүү бирдиктерге (токендерге) бөлүү табигый тилди изилдөө үчүн керектүү шарт. Эгерде тилдер так пунктуацияга ээ болгондо, токендештирүү маселе жаратмак эмес эле. Жөнөкөй программдык жабдыктар деле пунктуациялык белгилердин жардамы менен тексттерди сөздөргө бөлүштүрүп коймок. Бирок жашоодо тилдер мындай пунктуацияларга ээ болбогону үчүн токендештирүү маселеси жаралып келет.

Морфологиялык анализдин башка атайын маселеси **лемматизациялоо** болуп саналат, б.а. сөз колдонумдардын негизинде сөздүн баштапкы формасын алуу болуп эсептелет. Көптөгөн тилдерде сөз бир нече формада кездешүүсү мүмкүн. Мисалы '-сүйлөө' этиши 'сүйлөп', 'сүйлөгөн', 'сүйлөгөндө' жана 'сүйлөөдө' сыяктуу формаларында колдонулат. '-сүйлөө' негизги

сөзү сөздүктө лемма деп аталат. Лемматизациялоо бул сөздүн бир нече формасын топтоо процесси жана анализ учурунда бул топ бир сөз катары каралышы болуп саналат.

Лемматизация процессине окшош бирок айырмачылыгы бар **стемминг** процесси. Стеммингдин негизги максаты сөздүн негизин (стемасын) табуу. Лемматизациядан болгон айырмачылыгы катары стеммер сөздүн контекстин билбестен иштеп чыгат, ошондуктан сөздү дифференциалдап бөлүп чыга албайт. Ошону менен бирге тактыгы төмөн болсо дагы стеммерлерди колдонуу оңой жана иштеп чыгуу ылдамдыгы бийик. Мисалы "ал" токени үчүн "ага" леммасы дал келет, бирок стемминг үчүн мааниси жок. "чуркоо" леммасы "чуркады" токени үчүн базалык формасы жана бул окшоштук стемминг процессинде да лемматизациялоо процессинде да көрсөтүлөт.

Стемминг жана лемматизациялоо мисалы катары [The] [quick] [brown] [fox] [jumped] [over] [the] [lazy] [dogs] сөздөр берилген.

Стемминг боюнча [quick] [brown] [fox] [jump] [over] [lazy] [dog] сөздөр гана берилсе, бул сөздөрдүн леммасы катары [the] [quick] [brown] [fox] [jump] [over] [the] [lazy] [dog] сөздөрү болот.

Лемматизациялоо сөз түркүмдөрү менен байланышкан жана корпустагы сөздөрдү окшош лексемаларга чейин кыскартып берет. Лемматизация гана изилдөөчүлөргө лексеманын ар түрдүү варианттарын бөлүп изилдөөгө мүмкүнчүлүк берет.

Парсинг – лексемалардын (сөз, токен) анын формалдуу грамматикасы менен болгон тартиптүүлүк боюнча салыштыруу процесси. Жыйынтык катары көз карандылык багы (синтаксикалык

бак) келет. Чоң көлөмдөгү корпустар үчүн автоматтык синтаксикалык анализ (парсер) түзүү компьютердик лингвистика үчүн абдан маанилүү тармак болуп эсептелинет.

Корпустарды алдын ала иштеп чыгууда эки шартты карап чыгуу зарыл:

1. Текстти даярдоодо корпустун баалуулугу, кийинки кадамдары үчүн роль ойногон ар бир кадам, корпусту түзүүчүгө лингвистикалык чечимдерди кабыл алуусун шарттайт. Колдонуучу издегенин табуу үчүн бул чечимдерди билүүсү зарыл. Мисалы текстти сөздөргө бөлгөн адам New York же болбосо Baden Baden сөздөрүн бир сөз катары кабыл алышы же эки сөз катары кабыл алышы керекпи? Буларга көңүл буруусу керек.

2. Колдонуучуга текстти алдын ала иштеп чыгууда кандай иштер болуп өткөнүн, каталык ыктымалдыктары ачык айкын берилиши керек, себеби, ар түрдүү жаңылыштыктар корпустун негизинде алынган жыйынтыктардын тууралыгына шарт түзөт.

ҮЧҮНЧҮ БӨЛҮМ

ТЕКСТТЕРДИ ИШТЕП ЧЫГУУ ҮЧҮН КОЛДОНУЛГАН ПРОГРАММАЛАР

3.1. Компьютердик лингвистика

Компьютердик лингвистика 1954-жылдары Джорджтаундын эксперименти болуп жаткандапайда болгон. Ошол убакта компьютер 60 жөнөкөй сүйлөмдү орусчадан англисчеге которгон. Андан кийинки агым 60-жылдары жасалма интеллект идеяларынын эволюциясында пайда болушуна туура келет. 80-жылдары болсо компьютердик лингвистика бир аз токтоп, андан кийин токтоосуз өнүгө баштады. 90-жылдардын башында болсо корпустук лингвистика статистикалык методдор жана анализдер менен иштөөгө өттү.

Тексттерди иштеп чыгуунун төмөндөгүдөй деңгээлдери бар:

1. *Графематикалык анализ.* Массивден сүйлөмдү жана сөз формаларын(токен) бөлүп алуу.
2. *Морфологиялык анализ.* Сөз формаларын сөз түркүмдөргө бөлүп алуу.
3. *Синтаксистик анализ.* Сүйлөмдөгү сөздөрдүн синтаксистик байланышын жана сүйлөмдүн синтаксистик структурасынын түзүлүшүн ачып көрсөтүү.
4. *Семантикалык анализ.* Сөздөр жана синтаксистик группалар арасындагы семантикалык байланышты ачып көрсөтүү.

Жогорудагы анализдин баарын автоматтык түрдө жасап чыгыш үчүн текстти автоматтык түрдө иштеп чыгуу программалары керек. Алардын ичине конкорданс кирет.

3.2. Конкорданс

Конкорданс –сүйлөмдөгү сөздөрдүн жана анын булактарынын жыйындысынын тизмеси. Бул термин корпустук лингвистикада кеңири колдонулат. Бул жерде сөз же сөз формалары алфавиттик түрдө шилтемеси менен тизмеленет.

Компьютерлер жок кезде конкордансты түзүү жумушу оор болгондуктан чоң мааниге ээ болгон Вед, Библия, Куран же У. Шекспидин чыгармалары үчүн гана конкорданстар жазалып келген.

Конкорданс төмөнкү лингвистикалык маселелерди чечүүдө колдонулат:

- Бир сөздүн ар түрдүү колдонуу жолдорун салыштыруу,
- Маанилүү сөздөрдү анализдөө,
- Сөз жана сүйлөмдөрдүн колдонуу жыштыгын изилдөө,
- Фраза, идиомаларды издөө жана текшерүү,
- Терминдер жана башка которууга керек болгон сөздөрдү издөө,
- Сөздөрдүн тизмесин түзүү (чыгармада көп колдонулган сөздөр).

Конкорданстын үлгүлөрү:

- Ф.М. Достоевский публицистика сөздүк-конкордансы (<http://dostoevskii.karelia.ru>),

- Nave's Topical Bible тематикалык конкордансы

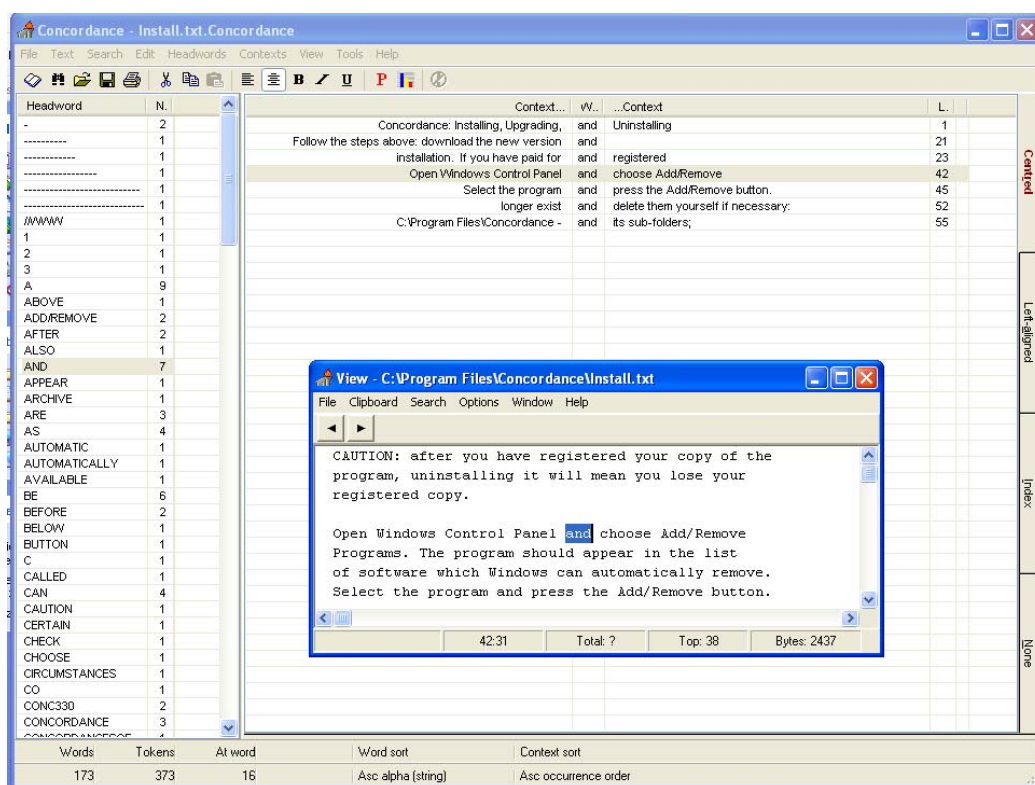
(<http://www.biblestudytools.com/concordances/naves-topical-bible/>)

- AntConc конкордансы

(<http://www.antlab.sci.waseda.ac.jp/software.html>)

- GlossaNet конкордансы(<http://glossa.fltr.ucl.ac.be/>)

- Unitex конкордансы (<http://www-igm.univ-mlv.fr/~unitex/>)



Сүрөт 1 Конкорданс программасынын иштөө чөйрөсү

Сүрөт 1 де конкорданстын иштөө чөйрөсүндө *.txt файлындагы тексттердин жыштыгы берилет. Мисалы бул жерде баардыгы 373 сөз жана 173 сөз формасы колдонулган. And сөзү 16-катарда турат жана ал бул текстте 7 жолу кайталанат. Ал эми оң жактагы терезеде ошол сөз колдонулган сүйлөмдөр жана ал сүйлөмдүн сабы берилген. Ошол сүйлөмдүн үстүнө бассак, ошол сүйлөм катышкан текст толугу

менен алдындагы терезечеде берилет. Бул программада ар кандай критерийлер менен издөө, алмаштыруу функциялары каралган. Бул тил илиминде абдан маанилүү. Ошол сөз эмне үчүн, эмне мааниде колдонуп жаткандыгын биле алышат.

3.3. Конкордансерлердин үлгүлөрү

Конкордансерлердин кээ бир үлгүлөрүнө төмөнкүлөр кирет:

- [TextSTAT - Simple Text Analyse Tool](#) – Ар түрдүү тил менен атайын сурак тилин (көп колдонулган сүрөттөө тили) колдонуп иштеген конкордансер.
- Орус, англис жана немис тилдер үчүн лингвистикалык издөө системасы (<http://aot.ru/>).

ТӨРТҮНЧҮ БӨЛҮМ

МАНАС ЭПОСУ ТУУРАЛУУ

4.1. “Манас” эпосу тууралуу кыскача маалымат.

«Манас» эпосу — кыргыз элинин улуттук сыймыгы. Кылымдарды карыткан руханий табылгалардын, кыргыздардын башынан кечирген окуялар жөнүндөгү маалыматтардын орчундууларын дээрлик бүт чогулткан алтын казынанын милдетин аткарып, элибиздин ыйык тутуп, укумдан-тукумга бөксөртпөй сактап, улам толуктап, байытып келе жаткан улуу мурасы, алтын казынасы.

«Манас» эпосу дүйнөдө тендешсиз зор көлөмдүү көркөм чыгарма. Манасчы Саякбай Карала уулунан жазылып алынган эпостун «Манас», «Семетей» жана «Сейтек» үч бөлүмүнүн жалпы көлөмү 500 миң сап ырды түзөт. Бул байыркы гректердин «Иллиадасы» менен «Одиссеясын» бирге кошкон көлөмүнөн 20 эседей көп. «Манастын» ошол бир эле варианты илимде узак убакыт эн көлөмдүү эпос делип келген инди элдеринин «Махабхаратасынан» эки жарым эсе чоң.

Жанрдык белгилери боюнча баатырдык эпос катары эсептелген бул чыгарманын мазмуну, берген маалыматтарынын көлөмү боюнча баатырдык эпостордун демейки алкагынан чыгып, андан алда канча тереңдиги, масштабдуулугу менен айырмаланат. Дүйнөгө белгилүү Ч. Валиханов, В. Радлов өндүү окумуштуулар XIX к. эле «Манас» эпосу аны жараткан кыргыз элинин тарыхын, философиясын, этнографиясын, тилин, көркөм өнөрүн, психологиясын, географиясын, медицинасын жалпы эле рухий жана социалдык

турмушунун ар кыл жактарын изилдеп үйрөнүүдөгү өзгөчө бай жана баалуу булак экендигинбаса белгилешкен.

«Манас» эпосу зор көлөмдүү, узак убакыт ооз жүзүндө гана жашап келгендиктен анын өзгөрүүсүз сакталган (канондук) тексти жок. Эпостун мазмуну, ыр саптары устаттан шакиртке, бир муундан экинчисине өткөндө, ал түгүл бир эле манасчы эпосту улам кайталап аткарганда өзгөрүүлөргө учураган. Ушундан келип эпостун түрдүү варианттары жаралган.

Азыркы кезде «Манас» эпосунун негизги бөлүмдөрүнүн кагаз бетине түшүрүлгөн 70тен ашуун варианты белгилүү. Эпостун улам жаңы үлгүлөрүн жыйноо иши дагы эле улантылууда. «Манастын» варианттарынын дээрлик бардыгында эпостогу негизги мазмунун, окуялардын нугун түзгөн салттык туруктуу эпизоддор сакталган. Бирок эпостун варианттарынын көлөмү, мазмундук, формалык касиеттери, көркөмдүк деңгээли бир кылка эмес. Бири-биринен орчундуу айырмачылыктарга ээ варианттар да, көп белгилери боюнча жакын турган, бир эле окуяны айрым өзгөчөлүктөр менен айткан варианттар да бар.

4.2. “Манас” эпосунун көлөмү.

“Манас” эпосунун көлөмү аябай чоң, мазмуну мол экенине дүйнөлүк илимпоздор XIX кылымдан баштап көңүл буруп, башка дүйнөлүк эпостор менен көлөмүн салыштырып, бой теңештире баштаган. 1856-жылы кыргыз жергесине келип “Көкөтөйдүн ашы” бөлүгү менен гана таанышкан Чокан Валиханов “Манас” эпосунун көлөмүнө кайрылып: “Кыргыздар “Манасты” үч түн бою укса да

айтып түгөтө алгыс, “Семетейге” да ошончо убакыт жетишсиз деп айтышат...” деген.

Ушул күндөгү маалыматтарга кулак төшөп көрсөк, Саякбай Каралаевдин айтымында кагаз бетине түшкөн “Манас” эпосунун бардык көлөмү 400 554 сап (караңыз, А.Жайнакова жазган “Манас” Баш сөз”, 2-бетинде. 1995-ж., Бишкек), Сагымбайдан бир гана “Манас” бөлүмү 180 378 сап (башка бөлүмдөрдү жазып алынбаган), Жусуп Мамайдан 300 000 сап кагаз бетине түшкөн. Бул үч манасчы “Манасты” эң көлөмдүү айткандар катарында саналат. Ал эми, ушул күнгө чейин “Манас” эпосунун түрдүү варианттарынын кагаз бетине түшүү көлөмү миллион сапка жеткирип маалымдаган басма сөздөр бар. Жүңгөдө 1950-жылдардан баштап төрт жолу “Манас” чогултуу жумушу колго алынып, бардыгы 600 000 сап “Манас” материалынын жазылгандыгын кабарлап келет (караңыз, Ху Чжен Хуа “Манас” изилдөөсү”. Май басма журналы, 1-сан. 1979-ж., 10-ай.).

“Манас” эпосунун көлөмдүүлүгүн, аны жазып алуусуна кеткен убакыт менен да бир ырастап көрсөтүүгө болот, алып айтсак, залкар манасчы Сагымбай Орозбак уулунан “Манастын” биринчи бөлүгүн 1922-жылы кагаз бетине түшүрүүнү баштап, нечен айлап, жылдап жазып, 1926-жылы төрт-беш жылда зорго жазып алуу иши байыган (албетте, аралыкта үзүлүп турса да). Ал эми залкар манасчы Саякбай Каралаевдин “Манас” эпосун жазып алуу иши 1931-жылы башталып, алды-кийин болуп, 1947-жылы байыган. Саякбайдан айттырып жазып алуу ишине 16 жыл кеткен. Ал эми залкар манасчы Жусуп Мамайдан “Манас” эпосун жазып алуу жумушу биринчи жолу 1961-жылы башталып, 1964-жылга чейин алты бөлүктү төрт жылда жазып алынган. Экинчи жолу 1979-жылдан 1982-жылга чейин дагы эки

бөлүктү толуктап жазууга үч жыл кеткен. 1982-жылдан 1985-жылга чейин башкалар жагынан кагаз бетине түшкөн өз айтымын (вариантын) өзү баштан аякка чейин көрүп, толуктап, өз колу менен кайра кагаз бетине түшүрүп, (кээ бир бөлүмүн) дагы үч жыл сарпталган. Ошентип, Жусуп Мамайдын “Манасы” кагаз бетине түшүрүлгөн убакыт 10 жыл болгон (аралыкта үзүлгөн убакыт эсептелбегенде).

“Манас” бөлүмү.

- 1.Сагымбай айтымында 180 378 сап;
- 2.Саякбай Каралаевдин айтымында 84513 сап;
3. Мамбет Чокморовдун айтымында 103090 сап;.
4. Жусуп Мамайдын айтымында 53280 сап.

“Семетей” бөлүмү.

- 1.Саякбай Каралаевдин айтымында 316157 сап;
2. Мамбет Чокморовдун айтымында 71609 сап;
3. Ш. Ырысмендеевдин айтымында 42338 сап;
4. Б. Сазановдун айтымында 67704 сап;
5. Жусуп Мамайдын айтымында 35246 сап;
6. Ыбырай Абдырахмановдун айтымында 23584 сап;
7. Эшмат Мамбетжусуптун айтымында 12152 сап.

“Сейтек” бөлүмү.

- 1.Саякбай Каралаевдин айтымында 84697 сап;
2. Жакшылык Сарык уулунун айтымында 27000 сап;
3. Мамбет Чокморовдун айтымында 23340 сап;

4. Жусуп Мамайдын айтымында 22590 сап;

5. Шапак Ырысмендеевдин айтымында 14718 сап.

Соңунда салыштырмалуу түрдө дүйнөгө даңазалуу эпостор – гректердин “Илиадасы” 15693 сап, “Одиссейя” 12110 сап экенин гана көрсөтсөк, көлөм салыштырмасы көрүнөт.

4.3. “Манас” эпосун изилдөө жана аны изилдөөчүлөр

«Манас» эпосун ар кандай аспектилер боюнча окумуштуулар изилдеп башташты. Алардын ичинде Валиханов, Родлов, Алмаши, Гүрсой сыяктуу чет элдик окумуштуулар да бар.

«Манас» эпосуна биринчи илимий пикир айткан адам Ч.Валиханов. Ал минтип жазган: «Манас – бир мезгилге келтирилген жана бир кишинин –Манас баатырдын тегерегине топтолгон бардык кыргыз мифтеринин, жомокторунун, санжыраларынын энциклопедиялык жыйнагы». Айрыкча «Манас» эпосун жыйноо, бастырып чыгаруу, изилдөө багытында орчундуу иштер иштелди. Кыргыз фольклорун изилдөөдө көрүнүктүү салым кошкондорго К.Рахматуллин, М.Богданова, Ж.Таштемиров, Б.Кебекова, С.Закироваларды да кошууга болот. Ошондой эле айрыкча «Манас» эпосунун түрдүү проблемаларын иликтөөдө зор салым кошуу менен орчундуу эмгектерди жазгандардын катарына: М.Ауэзов, С.Мусаев, Р.Кыдырбаева, Б.Юнусалиев, Р.Сарыпбеков, В.Жирмунский, С.Бегалиев, Э.Абдылдаевдер кирет.

4.5 Манас айтуучулар

I. Айткандары жазылбай калган манасчылар

1.1. Жайсаң

Белгилүү манасчы Сагымбай Орозбаковдун айтуусунда Жайсаң деген ырчы Манастын учурунда жашап, анын жортуулдарын даңктап ырдаган экен. Мисалы, ал боз үйдүн жабдылышын эле жарым күн айтчу экен. Чынында эле Манас эпосунун текстинде да Ырчы-уул деген баатыр эскерилет.

1.2. Токтогул ырчы

Мындан 500 жыл мурун жашаган Токтогул ырчы да Манас айткан деген сөз бар

1.3. Келдибек Барыбозов

Ал Көзкамандар окуясын эле бир жума айтчу экен. Келдибектин таланты тууралуу мындай легенда бар: ал Манас айта баштаганда күчтүү шамал согуп, жер титиреп, жайытта жүргөн мал айылга өзү келип калчу экен. Келдибек элдин эсинде табып катары да калыптыр. Кээ бир оорулуулар Манас уккандан кийин айыгып кетишчү экен.

1.4. Балык (Бекмурат) Кумаров

Анын өз аты Бекмурат болгон, а Балык – чыгармачыл псевдоними.

Балык өзүн Манастын чоролорунун бириндей сезип, өтө көтөрүңкү маанай менен ырдаган. Тилекке каршы, Балыктын айткан Манасы жана башка ырлары жазылып алынбагандыктан бизге жеткен жок. Балыктын уулу Найманбай да манасчы болгон, бирок ал кишинин да өмүр-баяны толук изилденген эмес.

1.5. Тыныбек Жапиев

Тыныбектин аткаруусундагыгы Манас да жазылып алынбаса да, анын Манас үчилтигин толук айткандыгы белгилүү .

1.6. Чоюке Өмүров

Чоюке манасты эң жакшы айткан устаттардын бири болгон. Ал Бакайдын окуясын айтып жатканда көпчүлүк угармандар ыйлашчу экен. Чоюкенин варианты да жазылбай калган

II. Айткандары жазылып алынган манасчылар

2.1. Сагымбай Орозбаков

1867-жылы Ысык-Көл областынын Кабырга айылында туулган. Атасы Орозбак саяк урусунан чыккан таланттуу сурнайчы болгон.

Жаш жана өспүрүм кездерин Нарын областындагы Кочкор айылында өткөргөн. Кыргызстандын аймактарын кыдырып жүрүп Манас айтып үй-бүлөсүн баккан.

Сагымбай – белгилүү манасчылар Балык, Келдибек, Чонбаш (Нармантай) жана анын окуучусу Тыныбектердин замандашы болгон жана алардын аткаруучулук чеберчилигин үйрөнгөн. Манасты 15 жашынан Чонбаш манасчыдан үйрөнө баштаган. Ага чейин тамаша ырларды, кошокторду чыгарып жана аткарып жүргөн.

"Манас", "Семетей" жана "Сейтек" үчилтигин толук айткан. С. Орозбаков Манас эпосунун эң көлөмдүү, сюжеттери толук вариантын жогорку көркөмдүк деңгээлде айткан. Анын вариантын окумуштуулар классикалык вариант катары кабыл алышат. С. Орозбаков 1930-жылы Кочкор районунда каза болгон. Сагымбайдын аткаруусунда үчилтиктин биринчи бөлүгү (Манас) толук жазылып алынган. Жазуу иши 1922-жылы Туркестан республикасынын Наркомпросунун тапшырмасы боюнча башталып, 1926-жылы аяктаган. Андан ары Сагымбайдын ооруп калгандыгына

байланыштуу жазуу улантылган эмес. Сагымбайдын Манасы 260 басма табактан ашып, кыргыз фольклорунун эң жакшы үлгүлөрүнүн бири болуп саналат.

2.2. Саякбай Каралаев

Бугу уруусунан чыккан бул манасчы 1894-жылы Ысык-Көл областынын Тоң районунда туулган. 1924-жылдары Саякбай Чоюке менен кездешип, ал айткан Манастын тексттерин үйрөнө баштайт. Манас тууралуу Саякбайга чоң энеси айтып берген. Ал киши өзү манас айтпаганы менен эпостун мазмунун жатка билчү. Саякбай "түш көргөндөн жети жылдан кийин" манас айта баштадым деп эскерет. Анын аткаруусунда үчилтиктин бардык бөлүктөрү (Манас, Семетей, Сейтек) жана Манастын урпактары тууралуу да бөлүктөр жазылып алынган.

БЕШИНЧИ БӨЛҮМ

МАНАС ЭПОСУНУН КОРПУСТАРЫН КОЛДОНУУ ЖАНА АНЫ ПРОГРАММА ЖАРДАМЫ МЕНЕН АНАЛИЗДӨӨ

5.1. Манас эпосунун корпустарында сөз формаларынын орточо кайталанышы.

Азыркы учурда сөздүк түзүү, грамматиканы өнүктүрүү сыяктуу баардык лингвистикалык изилдөөлөрдө лингвистиканын методдорун колдонбой жана маалыматтар базасын түзбөй иш жүзүнө ашыруу өтө татаал.

Ошондуктан корпустук лингвистика методдорунун ыкмаларынын жардамы менен кыргыз элинин эпос жанрдагы тексттерин изилдөөдө аны бөлүкчөлөргө бөлүп лингвистикалык корпус түзүлдү. Маалыматтык база катары төмөндө көрсөтүлгөн тексттер тиешелүү калыпта компьютерге жүктөлдү.

1. Бексултан Жакиев. Манас Кыргыздардын баатырдык эпосу.
2. Жайсаң Үмөт уулу Айткөл манас
3. "Манас" эпосу. Жусуп Мамайдын айтуусунда
4. "Манас" эпосу. Каралаев Саякбайдын айтуусунда
5. Манас биринчи бөлүк Жалпы редакцияланган – проф. Б.М.Юнусалиев
6. Манас экинчи бөлүк Жалпы редакцияланган – проф. Б.М.Юнусалиев
7. Семетей Жалпы редакцияланган – проф. Б.М.Юнусалиев
8. Семетей IЖусуп Мамайдын айтуусунда
9. Семетей IIЖусуп Мамайдын айтуусунда
- 10.Семетей IIIЖусуп Мамайдын айтуусунда

11."Семетей" эпосу. Каралаев Саякбайдын айтуусунда

12."Сейтек" эпосу. Каралаев Саякбайдын айтуусунда

Изилдөөдө текстте колдонулган сөз жана анын грамматикалык формаларын аныктоо үчүн жыштык сөздүк түзүлдү. Анын негизинде ар кандай статистикалык маалыматтар алынды.

Сөз формасынын орточо кайталанышы төмөнкү формула менен аныкталат.

$$F_{\text{орт}} = \frac{N}{L_{\text{с/ф}}}$$

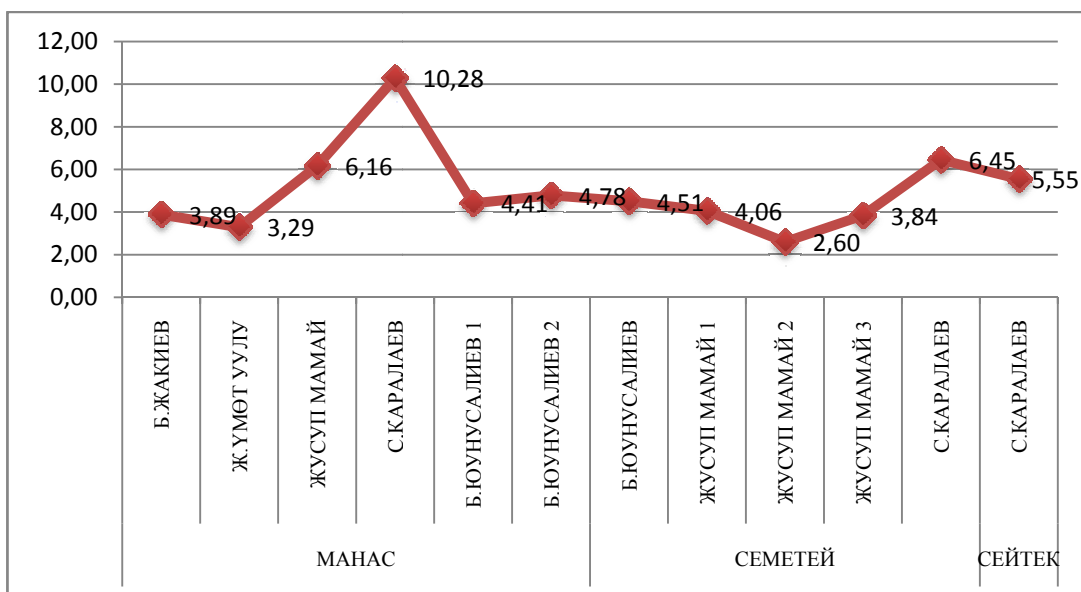
Бул жерде: $F_{\text{орт}}$ - текстте колдонулган сөз формаларынын орточо жыштыгы;

$L_{\text{с/ф}}$ - текстте колдонулган сөз формаларынын саны;

N – тексттин көлөмү.

Таблица 5.1.1. Сөз формаларынын орточо кайталанышы (толук текст)

КИТЕП	ВАРИАНТ	N	$L_{\text{с/ф}}$	$F_{\text{орт}}$
МАНАС	Б.ЖАКИЕВ	52195	13424	3,89
	Ж.ҮМӨТ УУЛУ	41436	12576	3,29
	ЖУСУП МАМАЙ	175317	28445	6,16
	С.КАРАЛАЕВ	256472	24954	10,28
	Б.ЮУНУСАЛИЕВ 1	72684	16479	4,41
	Б.ЮУНУСАЛИЕВ 2	72508	15162	4,78
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	71612	15894	4,51
	ЖУСУП МАМАЙ 1	59433	14637	4,06
	ЖУСУП МАМАЙ 2	13846	5322	2,60
	ЖУСУП МАМАЙ 3	39394	10260	3,84
	С.КАРАЛАЕВ	100564	15596	6,45
СЕЙТЕК	С.КАРАЛАЕВ	97378	17550	5,55



Бул берилиштерди жана графиктерди анализдеп карай турган болсок баардык вариантта С.Каралаевдин айтуусунда сөздөрдүн орточо кайталыныны көбүрөөк. Андан кийинки орунда Жусуп Мамай турат.

Бирок бул көрсөткүч боюнча анализ жүргүзүү үчүн бирдей көлөмдөгү тексттер салыштыруу керек. Ошондуктан изилдөөлөр ар кандай көлөмдөгү (10 миң, 20 миң, 30 миң, 40 миң, 70 миң, 100 миң сөз колдонулушундагы) тексттер менен жүргүзүлдү.

Таблица 5.1.2. Сөз формаларынын орточо кайталанышы (N=10-20-30 миң сөз)

КИТЕП	ВАРИАНТ	10000		20000		30000	
		L _{с/ф}	F _{орт}	L _{с/ф}	F _{орт}	L _{с/ф}	F _{орт}
МАНАС	Б.ЖАКИЕВ	4265	2,34	7066	2,83	9676	3,10
	Ж.ҮМӨТ УУЛУ	4826	2,07	8142	2,46	10762	2,79
	ЖУСУП МАМАЙ	4176	2,39	7018	2,85	9211	3,26
	С.КАРАЛАЕВ	3472	2,88	5541	3,61	7060	4,25
	Б.ЮУНУСАЛИЕВ 1	4309	2,32	6905	2,90	9166	3,27
	Б.ЮУНУСАЛИЕВ 2	3993	2,50	6788	2,95	8816	3,40
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	4277	2,34	6767	2,96	8989	3,34

	ЖУСУП МАМАЙ 1	4284	2,33	7194	2,78	9462	3,17
	ЖУСУП МАМАЙ 2	4278	2,34	5323	3,76	0	0,00
	ЖУСУП МАМАЙ 3	4200	2,38	6779	2,95	8630	3,48
	С.КАРАЛАЕВ	3661	2,73	5773	3,46	7239	4,14
СЕЙТЕК	С.КАРАЛАЕВ	4150	2,41	6651	3,01	8642	3,47
	<i>ОРТОЧО</i>	<i>4158</i>	<i>2,42</i>	<i>6662</i>	<i>3,04</i>	<i>8138</i>	<i>3,14</i>

10 миң сөз колдонулушу болгон текстте орточо эсеп менен 4158 сөз формасы колдонулган жана алар 2,42 жолу кайталанылган.

Ал эми 20 миң сөз колдонулушта болгон текстте орточо эсеп менен 6662 сөз формасы болуп 3,04 жолу кайталанса, сөз колдонулушу 30 миң болгон текстте орточо эсеп менен 8138 сөз формасы жана орточо кайталанышы 3,14 кө туура келет.

Таблицадагы көрсөткүчтөр төмөнкү график менен тастыкталат:

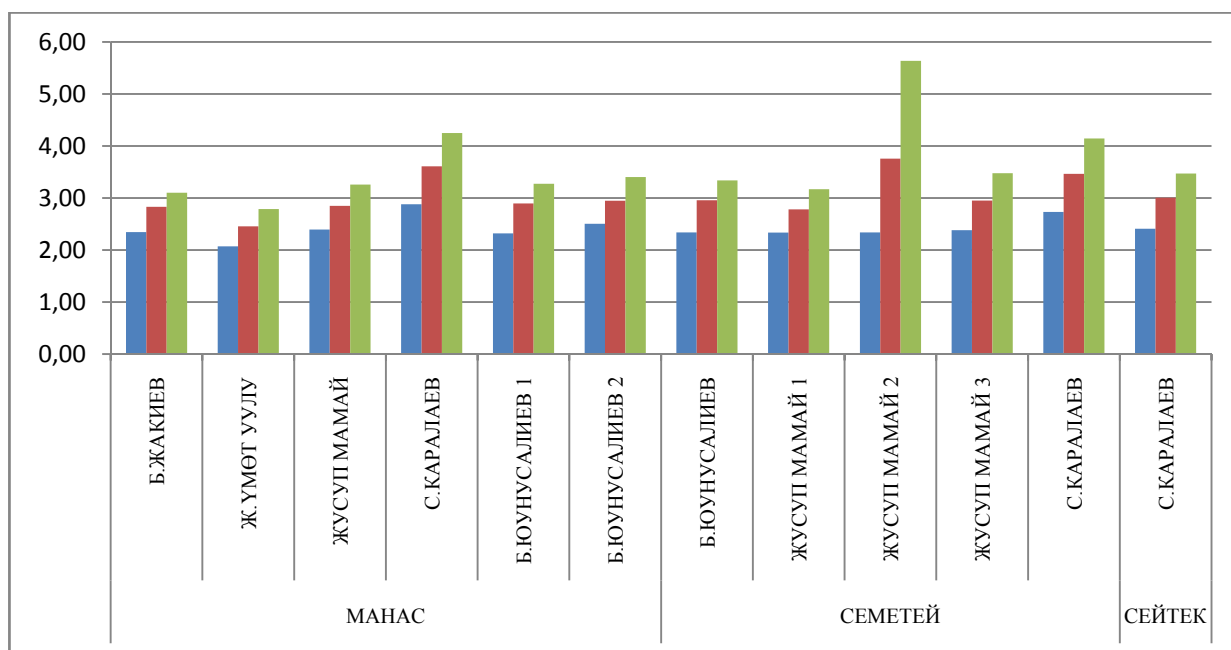


Таблица 5.1.3. Сөз формаларынын орточо кайталанышы (N=40-70-100 миң сөз)

КИТЕП	ВАРИАНТ	40000		70000		100000	
		L _c /ф	F _{орт}	L _c /ф	F _{орт}	L _c /ф	F _{орт}
МАНАС	Б.ЖАКИЕВ	11556	3,46	0	0,00	0	0,00
	Ж.ҮМӨТ УУЛУ	12343	3,24	0	0,00	0	0,00
	ЖУСУП МАМАЙ	11218	3,57	16246	4,31	21089	4,74
	С.КАРАЛАЕВ	8499	4,71	11705	5,98	14691	6,81
	Б.ЮУНУСАЛИЕВ 1	11284	3,54	16213	4,32	0	0,00
	Б.ЮУНУСАЛИЕВ 2	10883	3,68	14880	4,70	0	0,00
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	11059	3,62	15680	4,46	0	0,00
	ЖУСУП МАМАЙ 1	11369	3,52	14683	4,77	0	0,00
	ЖУСУП МАМАЙ 2	0	0,00	0	0,00	0	0,00
	ЖУСУП МАМАЙ 3	10282	3,89	0	0,00	0	0,00
	С.КАРАЛАЕВ	8897	4,50	12496	5,60	15638	6,39
СЕЙТЕК	С.КАРАЛАЕВ	10226	3,91	14189	4,93	17588	5,69
ОРТОЧО		9801	3,47	9674	3,26	5751	1,97

40 миң сөз колдонулушта болгон текстте орточо эсеп менен 9801 сөз формасы колдонулуп, алар 3,47 жолу кайталанат.

1-Таблицага назар салсак, изилдөөдө каралып жаткан сөздөр негизинен 40 миңдин тегерегинде болуп жатат. Ошондуктан 40 миң сөз колдонулушу болгон берилиштердин негизинде эң кичинекей көрсөткүч 3,24(Семетей эпосу Жусуп Мамай 2 вариантын эсепке албайбыз) Манас эпосу Жайсаң Үмөт уулунун вариантына туура келет. Демек Жайсаң Үмөт уулунун вариантында сөз жасоо мүмкүнчүлүгү жогору деп айтса болот. Андан кийинки көрсөткүчтө Манас эпосу Б.Жакиевдин вариантына туура келет. Бул көрсөткүчтөрдү төмөнкү графиктен салыштырмалуу көрсө болот.

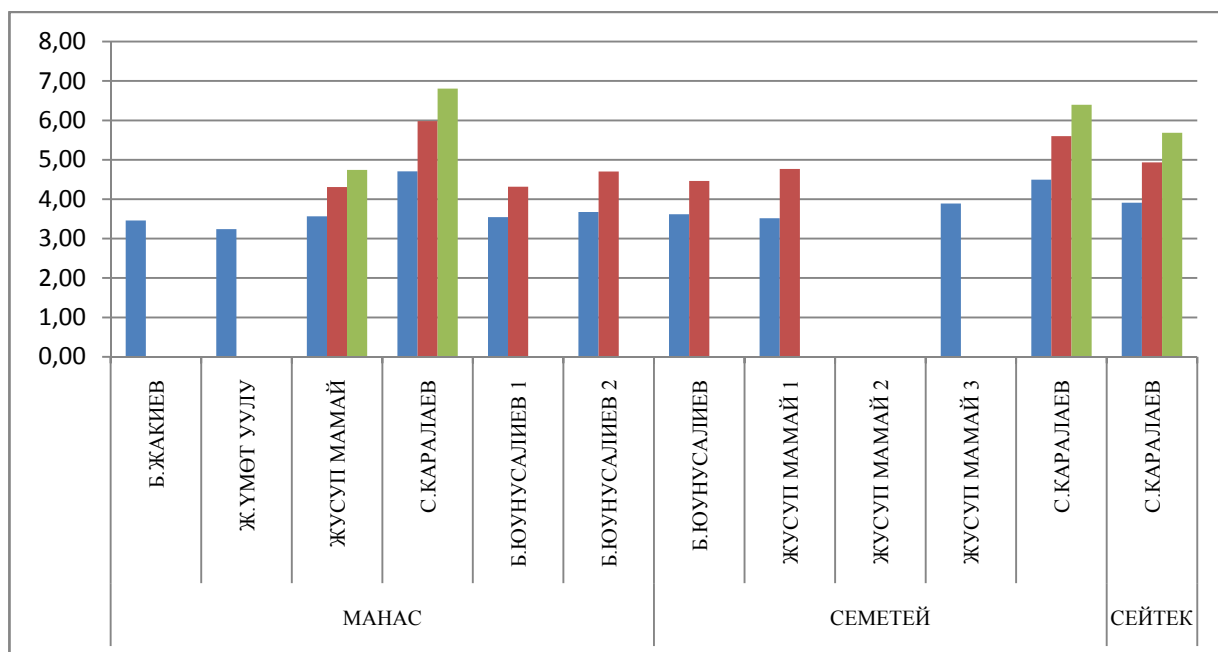


Таблица 5.1.4. Эпостордо тексттин көлөмүнө (N) карай сөз формаларынын орточо кайталанышынын (F) өсүшү

КИТЕП	ВАРИАНТ	10000		20000		30000	
		L _{с/ф}	F _{орт}	L _{с/ф}	F _{орт}	L _{с/ф}	F _{орт}
МАНАС	Б.ЖАКИЕВ	4265	2,34	7066	2,83	9676	3,10
	Ж.ҮМӨТ УУЛУ	4826	2,07	8142	2,46	10762	2,79
	ЖУСУП МАМАЙ	4176	2,39	7018	2,85	9211	3,26
	С.КАРАЛАЕВ	3472	2,88	5541	3,61	7060	4,25
	Б.ЮНУСАЛИЕВ 1	4309	2,32	6905	2,90	9166	3,27
	Б.ЮНУСАЛИЕВ 2	3993	2,50	6788	2,95	8816	3,40
СЕМЕТЕЙ	Б.ЮНУСАЛИЕВ	4277	2,34	6767	2,96	8989	3,34
	ЖУСУП МАМАЙ 1	4284	2,33	7194	2,78	9462	3,17
	ЖУСУП МАМАЙ 2	4278	2,34	5323	3,76	0	0,00
	ЖУСУП МАМАЙ 3	4200	2,38	6779	2,95	8630	3,48
	С.КАРАЛАЕВ	3661	2,73	5773	3,46	7239	4,14
СЕЙТЕК	С.КАРАЛАЕВ	4150	2,41	6651	3,01	8642	3,47
ОРТОЧО		4158	2,42	6662	3,04	8138	3,14

4-Таблицадын негизинде эпос жанрдагы тексттерде тексттин көлөмүнүн өсүшү сөз формаларынын орточо кайталанышы да тез өсөт деп айта албайбыз. С.Каралаевдин айтуусунда Манас эпосун

карасак $0,73 \rightarrow 0,64 \rightarrow 0,46 \rightarrow 1,27 \rightarrow 0,83$ ирээтинде болот. Ал эми Жусуп Мамайдын айтуусундагы Манас эпосун карасак $0,46 \rightarrow 0,41 \rightarrow 0,31 \rightarrow 0,74 \rightarrow 0,43$ ирээтинде. Башка варианттар дагы белгилүү темп менен өсүп же кемибейт. Муну менен биз эпостордогу сөз формасынын орточо кайталанышы тексттин көлөмүнө көз каранды эмес деп айта алабыз. Ал эми сөз формаларынын орточо кайталанышы тексттин көлөмүнө карай тез темп менен өсүп жатат.

5.2. Манас эпосунун корпустарында сейрек колдонулган сөз формаларынын текстти камтуу көрсөткүчү.

Статистикалык лексикографиянын маселелерин чечүүдө текстте эң жыш жана сейрек колдонгон сөз формаларын бөлүп алуу, андан соң алардын текстти камтуу өзгөчөлүктөрүн аныктоо талап кылынат. Бул маселени чечүү үчүн төмөнкү кадамдар аткарылышы керек.

1. Корпуста бир же эки жолу колдонулган сөз формаларын сейрек кездешүүчү сөз формалары деп, башка жыштыктагы сөз формаларын жыш кездешүүчү деп белгилейбиз.
2. Жыш колдонулган сөз формасынын текстти камтуу көрсөткүчү төмөнкү формула менен аныкталат.

$$= \left[1 - \frac{n(1,2)}{N} \right] * 100\%$$

Бул жерде : N – тексттин көлөмү;

$n(1,2)$ – бир жана эки жолу колдонулган сөз формаларынын жалпы саны.

**Таблица 5.2.1. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 10 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	☑
МАНАС	Б.ЖАКИЕВ	10000	4265	2895	67,88	3504	82,16	64,96
	Ж.ҮМӨТ УУЛУ	10000	4826	3303	68,44	4017	83,24	59,83
	ЖУСУП МАМАЙ	10000	4176	2780	66,57	3419	81,87	65,81
	С.КАРАЛАЕВ	10000	3472	2062	59,39	2625	75,60	73,75
	Б.ЮУНУСАЛИЕВ 1	10000	4309	2906	67,44	3547	82,32	64,53
	Б.ЮУНУСАЛИЕВ 2	10000	3993	2643	66,19	3232	80,94	67,68
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	10000	4277	2818	65,89	3478	81,32	65,22
	ЖУСУП МАМАЙ 1	10000	4284	2821	65,85	3485	81,35	65,15
	ЖУСУП МАМАЙ 2	10000	4278	2808	65,64	3458	80,83	65,42
	ЖУСУП МАМАЙ 3	10000	4200	2793	66,50	3395	80,83	66,05
	С.КАРАЛАЕВ	10000	3661	2223	60,72	2808	76,70	71,92
СЕЙТЕК	С.КАРАЛАЕВ	10000	4150	2659	64,07	3332	80,29	66,68

**Таблица 5.2.2. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 20 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	☑
МАНАС	Б.ЖАКИЕВ	20000	7066	4479	63,39	5572	78,86	72,14
	Ж.ҮМӨТ УУЛУ	20000	8142	5368	65,93	6608	81,16	66,96
	ЖУСУП МАМАЙ	20000	7018	4425	63,05	5541	78,95	72,295
	С.КАРАЛАЕВ	20000	5541	3098	55,91	4007	72,32	79,965
	Б.ЮУНУСАЛИЕВ 1	20000	6905	4382	63,46	5379	77,90	73,105
	Б.ЮУНУСАЛИЕВ 2	20000	6788	4256	62,70	5324	78,43	73,38
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	20000	6767	4115	60,81	5195	76,77	74,025
	ЖУСУП МАМАЙ 1	20000	7194	4525	62,90	5666	78,76	71,67
	ЖУСУП МАМАЙ 2	20000	5323	3406	63,99	4205	79,00	78,975
	ЖУСУП МАМАЙ 3	20000	6779	4182	61,69	5239	77,28	73,805
	С.КАРАЛАЕВ	20000	5773	3227	55,90	4180	72,41	79,1
СЕЙТЕК	С.КАРАЛАЕВ	20000	6651	4014	60,35	5080	76,38	74,6

**Таблица 5.2.3. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 30 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	Σ
МАНАС	Б.ЖАКИЕВ	30000	9676	6061	62,64	7480	77,30	75,0667
	Ж.ҮМӨТ УУЛУ	30000	10762	6883	63,96	8488	78,87	71,7067
	ЖУСУП МАМАЙ	30000	9211	5597	60,76	7085	76,92	76,3833
	С.КАРАЛАЕВ	30000	7060	3730	52,83	4895	69,33	83,6833
	Б.ЮУНУСАЛИЕВ 1	30000	9166	5639	61,52	7023	76,62	76,59
	Б.ЮУНУСАЛИЕВ 2	30000	8816	5244	59,48	6650	75,43	77,8333
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	30000	8989	5338	59,38	6780	75,43	77,4
	ЖУСУП МАМАЙ 1	30000	9462	5610	59,29	7163	75,70	76,1233
	ЖУСУП МАМАЙ 2	30000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 3	30000	8630	5116	59,28	6422	74,41	78,5933
	С.КАРАЛАЕВ	30000	7239	3838	53,02	5014	69,26	83,2867
СЕЙТЕК	С.КАРАЛАЕВ	30000	8642	4982	57,65	6423	74,32	78,59

**Таблица 5.2.4. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 40 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	Σ
МАНАС	Б.ЖАКИЕВ	40000	11556	6986	60,45	8741	75,64	78,1475
	Ж.ҮМӨТ УУЛУ	40000	12343	7064	57,23	9311	75,44	76,7225
	ЖУСУП МАМАЙ	40000	11218	6613	58,95	8428	75,13	78,93
	С.КАРАЛАЕВ	40000	8499	4435	52,18	5816	68,43	85,46
	Б.ЮУНУСАЛИЕВ 1	40000	11284	6878	60,95	8582	76,05	78,545
	Б.ЮУНУСАЛИЕВ 2	40000	10883	6374	58,57	8077	74,22	79,8075
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	40000	11059	6397	57,84	8168	73,86	79,58
	ЖУСУП МАМАЙ 1	40000	11369	6607	58,11	8446	74,29	78,885
	ЖУСУП МАМАЙ 2	40000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 3	40000	10282	5899	57,37	7526	73,20	81,185
	С.КАРАЛАЕВ	40000	8897	4667	52,46	6108	68,65	84,73
СЕЙТЕК	С.КАРАЛАЕВ	40000	10226	5704	55,78	7349	71,87	81,6275

**Таблица 5.2.5. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 70 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	█
МАНАС	Б.ЖАКИЕВ	70000	0	0	0,00	0	0,00	0
	Ж.ҮМӨТ УУЛУ	70000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ	70000	16246	9211	56,70	11792	72,58	83,1543
	С.КАРАЛАЕВ	70000	11705	5826	49,77	7628	65,17	89,1029
	Б.ЮУНУСАЛИЕВ 1	70000	16213	9160	56,50	11787	72,70	83,1614
	Б.ЮУНУСАЛИЕВ 2	70000	14880	8276	55,62	10587	71,15	84,8757
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	70000	15680	8680	55,36	11159	71,17	84,0586
	ЖУСУП МАМАЙ 1	70000	14683	8238	56,11	10588	72,11	84,8743
	ЖУСУП МАМАЙ 2	70000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 3	70000	0	0	0,00	0	0,00	0
	С.КАРАЛАЕВ	70000	12496	6241	49,94	8173	65,40	88,3243
	С.КАРАЛАЕВ	70000	12496	6241	49,94	8173	65,40	88,3243
СЕЙТЕК	С.КАРАЛАЕВ	70000	14189	7605	53,60	9859	69,48	85,9157

**Таблица 5.2.6. Эпостордо сейрек колдонулган сөз формаларынын
квантативдин мүнөздөмөсү (тексттин көлөмү 100 миң сөз)**

КИТЕП	ВАРИАНТ	N	L _{с/ф}	F ₁	F ₁ %	F _{1,2}	F _{1,2} %	█
МАНАС	Б.ЖАКИЕВ	100000	0	0	0,00	0	0,00	0
	Ж.ҮМӨТ УУЛУ	100000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ	100000	21089	11859	56,23	15104	71,62	84,896
	С.КАРАЛАЕВ	100000	14691	7073	48,15	9389	63,91	90,611
	Б.ЮУНУСАЛИЕВ 1	100000	0	0	0,00	0	0,00	0
	Б.ЮУНУСАЛИЕВ 2	100000	0	0	0,00	0	0,00	0
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	100000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 1	100000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 2	100000	0	0	0,00	0	0,00	0
	ЖУСУП МАМАЙ 3	100000	0	0	0,00	0	0,00	0
	С.КАРАЛАЕВ	100000	15638	7760	49,62	10065	64,36	89,935
	С.КАРАЛАЕВ	100000	15638	7760	49,62	10065	64,36	89,935
СЕЙТЕК	С.КАРАЛАЕВ	100000	17588	9352	53,17	12037	68,44	87,963

Жогорудагы таблицада бир ирет колдонулган сөз формаларынын саны F₁мамычасына, анын ошол тексттеги сөз формаларынын жалпы санына карата үлүшү F₁%мамычасына, бир

же эки ирет колдонулган сөз формаларынын жалпы саны $F_{1,2}$, ал эми анын үлүшү болсо $F_{1,2}\%$ мамычасында орун алат. Эң акырында жыш колдонулган сөз формаларынын үлүшү $F_{1,2}\%$ мамычасында берилет. Бул таблицада бир жана эки ирет колдонулган сөз формаларынын үлүшү азайып жатат. Демек тексттин көлөмү чоңойгон сайын сейрек жолуккан сөз формаларынын саны да азаюуда.

Таблица 5.2.7. Эпостордо сейрек колдонулган сөз формаларынын квантитативдин мүнөздөмөсү (толук текст үчүн)

КИТЕП	ВАРИАНТ	N	$L_{с/ф}$	F_1	$F_1\%$	$F_{1,2}$	$F_{1,2}\%$	$\bar{F}$
МАНАС	Б.ЖАКИЕВ	52195	13424	7877	58,68	9923	73,92	80,9886
	Ж.ҮМӨТ УУЛУ	41436	12576	7111	56,54	9461	75,23	77,1672
	ЖУСУП МАМАЙ	175317	28445	14964	52,61	19346	68,01	88,9651
	С.КАРАЛАЕВ	256472	24954	11388	45,64	15018	60,18	94,1444
	Б.ЮУНУСАЛИЕВ 1	72684	16479	9220	55,95	11899	72,21	83,6291
	Б.ЮУНУСАЛИЕВ 2	72508	15162	8366	55,18	10723	70,72	85,2113
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	71612	15894	8805	55,40	11311	71,17	84,2052
	ЖУСУП МАМАЙ 1	59433	14637	8199	56,02	10545	72,04	82,2573
	ЖУСУП МАМАЙ 2	13846	5322	3405	63,98	4204	78,99	69,6374
	ЖУСУП МАМАЙ 3	39394	10260	5876	57,27	7505	73,15	80,9489
	С.КАРАЛАЕВ	100564	15596	7689	49,30	10003	64,14	90,0531
СЕЙТЕК	С.КАРАЛАЕВ	97378	17550	9310	53,05	11996	68,35	87,681

Жогорудагы таблицада бир жолу колдонулган сөз формаларынын үлүшү орточо эсеп менен 54,97 үлүшүн түзсө, бир же эки жолу колдонулган сөз формаларынын үлүшү 70,68ди түзөт.

5.3. Тамгалардын айрым статистикасы.

**Таблица 5.3.1. Сөздүк ичиндеги тамгалардын сандык статистикасы
(Сөздүн башы)**

Алфавит	МАНАС						СЕМЕТЕЙ					СЕЙТЕК
	Б.ЖАКИЕВ	Ж.УМӨТ УУЛУ	ЖУСУП МАМАЙ	С.КАРАЛАЕВ	Б.ЮНУСАЛИЕВ 1	Б.ЮНУСАЛИЕВ 2	Б.ЮНУСАЛИЕВ	ЖУСУП МАМАЙ 1	ЖУСУП МАМАЙ 2	ЖУСУП МАМАЙ 3	С.КАРАЛАЕВ	С.КАРАЛАЕВ
А	6844	5262	23162	30485	9302	9326	8845	8649	1677	4641	11373	10481
Б	6934	5105	24489	35527	9868	9297	8700	7304	1828	4806	14036	13647
В	0	4	12	0	0	0	0	11	0	0	0	5
Г	33	168	74	317	87	70	161	337	88	384	297	75
Д	1951	1757	6693	7939	2668	2530	2800	1897	472	1282	2670	3148
Е	0	0	1	0	0	2	0	2	0	0	3	1
Ё	0	0	0	0	0	0	0	0	0	0	0	0
Ж	4983	3941	15757	23368	7227	6626	6811	5226	1292	3830	9391	9337
З	342	476	1120	1894	434	473	579	381	86	326	765	762
И	342	442	1482	1784	547	435	559	544	143	386	883	655
Й	3	17	57	34	5	5	5	6	2	3	12	17
К	13322	8377	41108	67695	18503	19516	17826	13889	3497	9695	25471	25448
Л	290	107	623	1585	346	371	459	203	52	219	675	610
М	2639	1458	8179	12292	3469	3185	2818	2366	656	1742	4395	3335
Н	475	469	1586	649	674	760	403	474	139	409	464	193
Ѓ	0	0	0	0	0	0		0	0	0	0	0
О	1315	975	4058	6856	1755	1775	1695	1322	347	855	2733	2396
Ө	296	306	539	514	365	335	382	281	100	222	346	401
П	98	171	328	73	133	95	94	106	33	62	76	77
Р	42	47	187	305	80	77	97	90	11	25	95	226
С	2522	2285	8070	12159	3535	3725	4085	3434	687	2329	5074	5021
Т	3926	4656	14132	20585	5663	5599	6208	4876	993	3056	8869	8597
У	915	1093	3521	4617	1305	1171	1341	1126	342	839	1975	2052
Ү	127	142	220	189	134	110	148	102	49	88	127	123
Ф	2	1	0	0	0	0	0	0	0	0	0	0
Х	4	331	887	0	0	0	1	349	57	254	1	0
Ц	0	2	0	0	0	0	0	0	0	0	0	0
Ч	1772	1238	6913	10219	2559	2927	3037	2576	492	1332	4162	4316
Ш	406	560	1733	2193	563	565	631	533	120	329	842	963
Щ	0	0	0	0	0	0	0	0	0	0	0	1
Ъ	0	0	0	0	0	0	0	0	0	0	0	0
Ы	216	289	770	1077	305	278	290	239	40	157	622	405
Ь	0	0	0	0	0	0	0	0	0	0	0	0
Э	2440	1755	8667	13388	3142	3256	3551	2988	632	1899	5062	5136
Ю	0	1	0	0	0	0	0	0	0	0	0	0
Я	0	0	0	0	0	0	0	0	0	0	0	0

Жогрудагы таблицада көрүнүп тургандай сөздүн башында В, Е, Ё, Ф, Ц, Щ, Ъ, Ь, Ю, Я тамгаларынын колдонулушу жокко эсе. Ал эми К тамгасынын сөздүн башында колдонулушу эң көп көрсөткүчтү көрсөтүп 67695 турат. Андан кийинки орундарда Б жана А тамгалары турат.

Х тамгасынын сөздүн башында колдонулушу Жайсаң Үмөт уулу жана Жусуп Мамайдын айтуусунда көп кездешет. Бул сөздөрдү жыштык сөздүктөн карай турган болсо бул сөздөрдүн көпчүлүгү “хан” сөзүнүн сөз формалары болуп саналат.

**Таблица 5.3.2. Сөздүк ичиндеги тамгалардын сандык статистикасы
(Сөздүн ортосу)**

Алфавит	МАНАС						СЕМЕТЕЙ					СЕЙТЕК
	Б.ЖАКИЕВ	Ж.ҮМӨГҮҮЛҮ	ЖУСУП МАМАЙ	С.КАРАЛАЕВ	Б.ЮНУСАЛИЕВ 1	Б.ЮНУСАЛИЕВ 2	Б.ЮНУСАЛИЕВ	ЖУСУП МАМАЙ 1	ЖУСУП МАМАЙ 2	ЖУСУП МАМАЙ 3	С.КАРАЛАЕВ	С.КАРАЛАЕВ
А	16197	11312	52894	79829	22091	22484	20337	17143	4081	11687	30195	28607
Б	3426	2363	12176	15942	4238	5298	4540	3869	855	2522	5623	5646
В	3	1	2	4	6	5	2	2	0	0	0	10
Г	7215	6598	23951	35806	10134	10655	10180	8409	1920	5418	14091	13676
Д	8544	6971	29019	40825	11285	11179	12320	10097	2534	6803	15636	14981
Е	7853	7031	28950	38345	11049	10746	12237	10058	2272	6816	16473	14877
Ё	69	30	1	267	74	98	92	10	2	7	82	82
Ж	638	405	1948	3956	539	800	725	603	130	354	1004	1127
З	2152	1971	5962	10238	2986	2799	2645	2166	619	1609	4085	4231
И	4044	4559	14697	20311	5612	5109	5940	4578	1111	3058	8633	7651
Й	5355	3584	18865	25260	7350	7255	7680	6125	1328	3879	9934	9643
К	5571	4399	18280	25335	7298	7172	7421	6481	1368	3698	10418	9126
Л	12951	9184	42907	64713	18248	18086	17656	14597	3560	9732	23640	25211
М	3146	2819	10559	15581	4454	5021	5501	4095	886	2800	7018	5732
Н	7525	5130	21333	33006	9497	9305	9421	7893	1881	5105	13612	11754
Ѓ	1453	1323	3368	2836	1831	1495	1722	1524	406	1075	1634	1963
О	4304	3268	13838	21177	6052	6188	6019	4868	1014	3059	7850	8643
Ө	1553	1656	3270	2770	1872	1789	1926	1487	544	1104	1917	2085
П	1574	1382	5474	7475	2421	2099	1764	1457	365	969	2595	3204
Р	11389	9408	39649	56309	15316	16306	15861	13241	3337	8946	21677	22338
С	3698	3060	11083	16642	4885	4812	5074	3997	834	2595	6225	6442
Т	8085	6116	26765	35827	10496	10798	10337	8889	2121	5797	13814	13528
У	5072	4318	17886	26094	7042	7397	7217	5426	1275	3576	10081	10760
Ү	1557	1963	3261	2822	1933	1756	1941	1542	567	1145	1920	2144
Ф	0	7	0	0	0	0	0	0	0	0	0	1
Х	2	30	168	9	5	11	3	10	0	2	12	17
Ц	0	2	0	0	0	0	0	0	0	0	4	0
Ч	1904	1792	7030	9495	2690	2598	2972	2796	630	1857	3710	3800
Ш	2851	2718	11014	13202	4195	4400	4067	3518	797	2204	4643	5008
Щ	1	0	0	0	1	1	0	0	0	0	1	0
Ъ	0	0	0	0	1	1	0	0	0	0	0	0
Ы	10912	7832	35743	54091	15222	15200	14240	12241	2770	7733	19982	19924
Ь	0	56	0	0	0	0	0	0	0	0	0	0
Э	235	261	308	1701	141	154	103	131	19	99	260	176
Ю	309	275	1	1328	462	423	379	333	114	251	501	549
Я	292	474	0	1633	580	435	514	311	66	229	804	929

Сөздүн ортосунда А, Р, Ы тамгаларынын колдонулушу эң көп
көрсөткүчтү көрсөтүп турат. Ы тамгасы сөздүн ортосунда Жайсаң

Үмөт уулунун айтуусунда көп колдонулган. Ал сөздөрдүн көпчүлүгү жердин аттары: Вайдурья, Гай-Юань-Тхай, Гуйянь-Ган, Гуйянь-Гань, Сянь-Фань, Тянь-Ян, Хэнь-Ян, Цуань-Чжу, Юань-Тхай сыяктуу сөздөр болуп саналат. Жусуп Мамай X тамгасын башка манасчыларга салыштырмалуу көбүрөөк колдонгон.

**Таблица 5.3.3. Сөздүк ичиндеги тамгалардын сандык статистикасы
(Сөздүн аягы)**

Алфавит	МАНАС						СЕМЕТЕЙ					СЕЙТЕК
	Б.ЖАКИЕВ	Ж.ҮМӨТҮҮЛҮ	ЖУСУП МАМАЙ	С.КАРАЛАЕВ	Б.ЮНУСАЛИЕВ 1	Б.ЮНУСАЛИЕВ 2	Б.ЮНУСАЛИЕВ	ЖУСУП МАМАЙ 1	ЖУСУП МАМАЙ 2	ЖУСУП МАМАЙ 3	С.КАРАЛАЕВ	С.КАРАЛАЕВ
А	4622	3486	14660	21012	5852	6411	5470	4857	1274	3282	7889	7225
Б	23	14	67	64	24	29	19	11	3	17	35	35
В	0	0	0	0	0	0	0	0	0	0	0	1
Г	129	200	555	636	191	163	232	171	46	106	335	418
Д	220	207	763	1047	267	322	333	201	87	218	527	519
Е	1523	1542	6299	8142	2070	2191	2518	2007	512	1387	3429	2953
Ё	22	5	0	53	12	28	6	2	0	0	33	15
Ж	14	2	52	112	23	27	16	10	7	12	37	12
З	934	908	2761	3800	1321	980	915	771	206	704	1503	1962
И	2192	1340	7081	9691	2790	2743	2427	2389	576	1604	3624	2866
Й	2459	1518	8728	13618	3196	3784	4383	3287	773	2292	5812	5464
К	2294	1855	7718	13447	3537	3623	3483	2531	567	1660	5069	4900
Л	1397	1618	5103	7801	2015	1887	1830	1520	402	1072	2952	3393
М	1384	2592	4346	7227	1953	1875	2653	1793	324	1113	3752	2967
Н	10547	10041	34929	56105	15031	15604	15705	12688	2788	8229	21722	20187
ң	550	528	1200	1104	724	534	636	600	158	444	688	791
О	902	654	2564	4699	1248	1232	1840	1553	331	1119	2085	1818
Ө	274	334	589	500	336	338	367	268	124	221	346	396
П	7604	4697	28860	37449	11125	10463	11259	9126	2189	6156	14604	16223
Р	3365	2228	11670	14929	5040	4777	3488	3186	819	1992	5250	6211
С	936	370	2072	3877	1161	997	780	629	108	365	1140	971
Т	1614	1424	5889	7835	2225	2413	2223	1989	379	1060	2731	2415
У	1936	1432	6139	9755	2759	2666	2599	2316	556	1472	4134	4124
Ү	440	419	884	799	547	517	525	423	166	321	551	580
Ф	0	0	0	2	0	0	0	0	0	0	0	0
Х	0	1	0	0	0	0	0	0	0	0	0	0
Ц	0	0	0	0	0	0	0	0	0	0	0	0
Ч	243	240	986	1166	444	281	288	238	65	227	417	592
Ш	540	466	2180	2453	771	775	728	679	132	388	1095	1004
Щ	0	0	0	0	0	0	0	0	0	0	0	0
Ъ	0	0	0	0	0	0	0	0	0	0	0	0
Ы	4330	2073	13199	17291	5533	5310	4247	4079	1042	2741	6351	5808
Ь	0	20	0	0	0	0	0	0	0	0	0	0
Э	56	34	150	258	77	59	68	35	11	39	156	110
Ю	31	21	0	116	69	49	39	45	9	34	63	52
Я	49	25	0	245	63	81	77	35	22	48	152	89

көрсөтүп турат.

5.4. Сөздөрдүн узундугу боюнча маалыматтар.

Таблица 5.4.1. Сөздүн узундугу боюнча маалыматтар

МАНАС										СЕМЕТЕЙ										СЕЙТЕК			
Б.ЖАКИЕВ		ЖҮМӨТ УУЛУ		ЖУСУП МАМАЙ		С.КАРАЛАЕВ		Б.ЮНУСАЛИЕВ 1		Б.ЮНУСАЛИЕВ 2		Б.ЮНУСАЛИЕВ		ЖУСУП МАМАЙ 1		ЖУСУП МАМАЙ 2		ЖУСУП МАМАЙ 3		С.КАРАЛАЕВ		С.КАРАЛАЕВ	
Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны	Сөз узундугу	Анын саны
19	1	21	1	23	1	20	2	16	1	19	1	19	2	17	2	17	1	14	2	24	1	19	1
17	1	19	9	21	1	18	3	15	2	17	2	16	2	14	2	13	5	13	8	20	1	18	1
15	2	18	4	20	2	17	3	14	4	16	1	15	1	13	18	12	34	12	67	17	6	16	4
14	7	17	1	18	1	16	11	13	41	15	1	14	5	12	86	11	122	11	335	16	5	15	2
13	29	16	7	17	1	15	15	12	211	14	7	13	45	11	463	10	224	10	684	15	5	14	9
12	167	15	6	16	1	14	26	11	631	13	35	12	182	10	1153	9	677	9	1773	14	5	13	65
11	458	14	4	15	9	13	193	10	1315	12	241	11	735	9	2912	8	1681	8	4664	13	43	12	294
10	1081	13	39	14	7	12	785	9	3058	11	710	10	1593	8	7187	7	2169	7	5954	12	204	11	1008
9	2376	12	283	13	78	11	2727	8	8255	10	1628	9	3706	7	9414	6	2146	6	6240	11	1005	10	2115
8	6159	11	972	12	420	10	5272	7	10817	9	3738	8	8235	6	9001	5	2994	5	8641	10	1753	9	4609
7	7752	10	1862	11	1682	9	10966	6	11915	8	8648	7	11241	5	13243	4	1579	4	4427	9	4212	8	10787
6	8693	9	2370	10	3047	8	29191	5	16899	7	10134	6	11264	4	6656	3	1814	3	5428	8	10818	7	14691
5	11871	8	3715	9	7970	7	34634	4	7737	6	11879	5	15490	3	7504	2	395	2	1150	7	15160	6	17057
4	5427	7	4255	8	21868	6	42934	3	9609	5	15966	4	7536	2	1779	1	5	1	21	6	16512	5	20205
3	6566	6	6953	7	25032	5	58188	2	2136	4	7769	3	9424	1	13					5	22339	4	10094
2	1576	5	9648	6	27000	4	29037	1	53	3	9413	2	2100							4	11593	3	12850
1	28	4	4865	5	40375	3	34580			2	2298	1	51							3	13866	2	3514
		3	5031	4	19896	2	7810			1	37									2	2940	1	72
		2	1278	3	22828	1	95													1	96		
		1	133	2	5051																		
				1	47																		

Жогорудагы таблицада көрсөтүлгөндөй эң көп 24 тамгага чейинки узундуктагы сөздөр бар. Таблицада көрүнүп тургандай сөз узундугу 13төн чоң болгондор аз кездешүүчү сөздөр болуп саналат. Төмөндө ар бир вариант үчүн сөз узундугу боюнча маалыматтар көрсөтүлгөн.

Таблица 5.4.2. Баардык вариант үчүн текст ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу

КИТЕП	ВАРИАНТ	СӨЗ	УЗУНДУГУ
МАНАС	Ж.МАМАЙ	Чаалыкпастыктын	15
МАНАС	ЖАЙСАН	тынчсызданткан	14
МАНАС	ЖАЙСАН	жийиркендирген	14
МАНАС	Б.ЮУНУСАЛИЕВ_1	калмаккытайдан	14
МАНАС	Б.ЮУНУСАЛИЕВ_1	КӨЗКАМАНДАРДЫН	14
МАНАС	Б.ЮУНУСАЛИЕВ_1	Айкөлдү кармап	14
МАНАС	Б.ЮУНУСАЛИЕВ_2	чыктыдалдайтып	14
МАНАС	Б.ЮУНУСАЛИЕВ_2	аманындакөрсүн	14
МАНАС	Б.ЮУНУСАЛИЕВ_1	колуң көргөндө	14
МАНАС	ЖАЙСАН	тынчсызданткан	14
МАНАС	ЖАЙСАН	мурдагысындай	13
МАНАС	ЖАЙСАН	жымжырттыктан	13
МАНАС	ЖАЙСАН	тынчсызданбай	13
МАНАС	Ж.МАМАЙ	кылычнайзасын	13
МАНАС	С.КАРАЛАЕВ	Бабадыйкандын	13
МАНАС	С.КАРАЛАЕВ	азуулулардын	13
МАНАС	С.КАРАЛАЕВ	Толгонтпостон	13
МАНАС	С.КАРАЛАЕВ	Кайкайтпастан	13
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	Тулпарларынын	13
МАНАС	Ж.МАМАЙ	Корккондуктун	13
МАНАС	Б.ЖАКИЕВ	Кабыландардан	13
МАНАС	ЖАЙСАН	болжолдоорсуң	13
МАНАС	Б.ЮУНУСАЛИЕВ_2	тердеткенинде	13
МАНАС	Б.ЮУНУСАЛИЕВ_1	Тарткандардын	13
МАНАС	Ж.МАМАЙ	Кыркжаңжүндүн	13
СЕМЕТЕЙ	Ж.МАМАЙ_2	БААТЫРЛАРЫНАН	13
СЕМЕТЕЙ	Ж.МАМАЙ_3	Таарынттырбай	13
СЕЙТЕК	С.КАРАЛАЕВ	Каарлангандай	13
МАНАС	Б.ЮУНУСАЛИЕВ_2	Арыстандардын	13
МАНАС	ЖАЙСАН	тоскоолдуксуз	13

**Таблица 5.4.3. Манас эпосу Б.Жакиевдин вариантындагы текст ичиндеги
сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Кабыландардан	13
2	Кабыландардын	13
3	пейилчечилдик	13
4	тердеткенинде	13
5	Айттырбастан	12
6	Арстандардын	12
7	БААТЫРЛАРДЫН	12
8	Баатырлардын	12
9	Баатырсынган	12
10	Бакдөөлөттүн	12
11	Билгендерден	12
12	Билдирбестен	12
13	Биртуугансың	12
14	Бозкертиктин	12
15	Буудайгүлдөн	12
16	Бээжиндиктин	12
17	Жайсаңбайдын	12
18	Жалгыздыктан	12
19	Жергенттикке	12
20	Жеткиргейсиң	12
21	Жолборсундун	12
22	Жолдоштордун	12
23	Жортуулчудай	12
24	Жумшаргансып	12
25	Жылгындуунун	12
26	Кабыланыңдын	12
27	Кайраттуусун	12
28	кайрылбастан	12
29	Кайтараармын	12
30	Калмактардын	12
31	Калтырбастан	12
32	Калтырбастан	12
33	Камбарбоздун	12
34	Каңгайлардын	12
35	Каңгайлыктын	12

**Таблица 5.4.4. Манас эпосу Ж.Үмөт уулунун вариантындагы текст ичиндеги
сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	жийиркендирген	14
2	тынчсызданткан	14
3	болжолдоорсуң	13
4	жымжырттыктан	13
5	мурдагысындай	13
6	тоскоолдуксуз	13
7	тынчсызданбай	13
8	баатырлардын	12
9	баатырлыктын	12
10	байкатпастан	12
11	байкатпастан	12
12	байкатпастан	12
13	байкоосуздан	12
14	байкоосуздан	12
15	бараткансыйт	12
16	баркталбаган	12
17	бастырбастан	12
18	Баткактуунун	12
19	бейжайландым	12
20	бейтааныштан	12
21	бейтааныштын	12
22	билбептирмин	12
23	билгендерден	12
24	билгендиктен	12
25	билгизбестен	12
26	билдирбестен	12
27	биртуугандын	12
28	бөксөртпөсөң	12
29	буйдалбастан	12
30	булганбастан	12
31	бүркөлмөктөн	12
32	бүткөндөрдөн	12
33	буудандардын	12
34	буурчактаган	12
35	дартмандууга	12

**Таблица 5.4.5. Манас эпосу Жусуп Мамайдын вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Чаалыкпастыктын	15
2	Башкарганынан	13
3	Буурултокойдо	13
4	Буурултокойду	13
5	Калксыздыктан	13
6	Камдалганынан	13
7	Карабөктөрдү	13
8	Корккондуктун	13
9	Кызылмойнокко	13
10	кылычнайзасын	13
11	Кыркжаңжүндүн	13
12	салгылашуулар	13
13	Айрылганынан	12
14	Айткандардын	12
15	Айттырбастан	12
16	Албууттанган	12
17	Арстандардын	12
18	Астакбыралла	12
19	Аттанганынан	12
20	Аттанганынан	12
21	баатырлардын	12
22	Баатырлыктан	12
23	Баатырлыктын	12
24	Баатырсынган	12
25	Багынганынан	12
26	Байгамбардай	12
27	Байдербиштин	12
28	Бакбурчундун	12
29	Бакдөөлөттүн	12
30	Балбандардан	12
31	Балбандыктан	12
32	Балбандыктын	12
33	Барктууланып	12
34	бейжиндиктин	12
35	Белдүүсүнгөн	12

**Таблица 5.4.6. Манас эпосу С.Каралаевдын вариантындагы текст ичиндеги
сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	азуулуулардын	13
2	Бабадыйкандын	13
3	Кайкайтпастан	13
4	Толгонтпостон	13
5	Тыргооттордон	13
6	Тыргооттордун	13
7	Айткандардын	12
8	Арсландардан	12
9	Баатырдырсың	12
10	Баатырлардан	12
11	Баатырлардын	12
12	Баатырсынган	12
13	Байгамбардай	12
14	Байгамбардын	12
15	Бакдөөлөттөй	12
16	Балбандардын	12
17	Балбансынган	12
18	Баргандардын	12
19	бастырбастык	12
20	Баткактуунун	12
21	Бейпайлашкан	12
22	Беттешкендин	12
23	Билгендерден	12
24	Билгичтерден	12
25	Билгичтердин	12
26	Билдирбестен	12
27	Бозкертиктин	12
28	Буйдалбастан	12
29	Булкунгандан	12
30	Буудайкандын	12
31	Буудандардын	12
32	Бууруксунган	12
33	Былкылдактын	12
34	Былкылдатпай	12
35	Бээжиндиктей	12

**Таблица 5.4.7. Манас эпосу Б.Юунусалиев 1 вариантындагы текст ичиндеги
сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	КӨЗКАМАНДАРДЫН	14
2	Тарткандардын	13
3	Арыстандыгың	12
4	Арыстанындын	12
5	Арыстанындын	12
6	Арыстаныңдын	12
7	Баатырлардын	12
8	Баатырсынган	12
9	Бакбурчундук	12
10	Бакдөөлөттүн	12
11	Балбандардын	12
12	Баргандардан	12
13	Бейбаштардын	12
14	Беттешкенден	12
15	билгендердин	12
16	Билгичтерден	12
17	Буудайбектен	12
18	Буудайбектин	12
19	Быркылдактын	12
20	Бытпылдыктап	12
21	Дөрбөндөрдүн	12
22	Жайсаңбайдын	12
23	Жайсаңбайдын	12
24	Жалгыздыктан	12
25	Желтийбестин	12
26	Жолборсундун	12
27	Жылаандардын	12
28	Жылгындуунун	12
29	Жылтыркандуу	12
30	Кабыланындын	12
31	Кабыланыңдын	12
32	Кайраттуулар	12
33	Кайраттуунун	12
34	Кайраттуусун	12
35	калмакманжуу	12

**Таблица 5.4.8. Манас эпосу Б.Юунусалиев 2 вариантындагы текст ичиндеги
сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Арыстандардын	13
2	тердеткенинде	13
3	Айттырбастан	12
4	Арыстандарың	12
5	Арыстанымдын	12
6	Арыстанындын	12
7	Баатырлардын	12
8	Баатырсынган	12
9	Баатырсынган	12
10	Балбандардын	12
11	Балбандардын	12
12	Билгендерден	12
13	Билгичтерден	12
14	Билдирбестен	12
15	Бозкертиктин	12
16	Буйдалбастан	12
17	Булкунгандан	12
18	Бээжиндиктин	12
19	Дардаңдардан	12
20	Дарттуусунун	12
21	Дөрбөндөрдүн	12
22	Жалгыздыктын	12
23	жембаштыктай	12
24	Жоомарттыгын	12
25	Жылгындуунун	12
26	Кабыландыгың	12
27	Кабыланымдын	12
28	Кабыланымдын	12
29	Кабыланыңдын	12
30	кайрандардын	12
31	Кайрылбастан	12
32	Калдайлардын	12
33	Калмактардын	12
34	Калтырбастан	12
35	Каңгайлардан	12

**Таблица 5.4.9. Семетей эпосу Б.Юунусалиевдинвариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Тулпарларынын	13
2	Анткорсунган	12
3	Баатырлардын	12
4	Баатырсынган	12
5	Баатырсынган	12
6	Байкуштардын	12
7	Бакдөөлөттүн	12
8	Балбандардын	12
9	Балтыркандын	12
10	Баткактуунун	12
11	Беттешпеймин	12
12	Билбестиктен	12
13	Билдирбеймин	12
14	Билдирбестен	12
15	Бозбуудандан	12
16	Буудандардан	12
17	Буудандардын	12
18	Доолаштырбай	12
19	Жайсандардын	12
20	Жансыздардан	12
21	Желдеңбайлап	12
22	Жоолашкандар	12
23	Жоолаштырбай	12
24	Жуукурланган	12
25	Жылтыркандуу	12
26	Кайгыргансып	12
27	Кайраттуубуз	12
28	Кайрылбастан	12
29	Кайрылгансып	12
30	Калкандардын	12
31	Калмактардын	12
32	Калтырбаймын	12
33	Калчуулардын	12
34	Кандайлыктан	12
35	Кармашкандан	12

**Таблица 5.4.10. Семетей эпосу Жусуп Мамай1 вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Арстандардын	12
2	Баатырсынган	12
3	Байлангандай	12
4	Балтанчандар	12
5	Баргандардан	12
6	Билдирбестен	12
7	Булдурсундун	12
8	Дасторкондун	12
9	Жезкемпирдин	12
10	Жоомарттыгын	12
11	Кабыландыгың	12
12	Кайгырбастан	12
13	Калмактардын	12
14	Калтырбастан	12
15	Каңгайлардан	12
16	Каңгайлардын	12
17	Канкорлордун	12
18	Караандатпай	12
19	Карматпастан	12
20	Карсылдашкан	12
21	Карткүрөңдөй	12
22	Карткүрөңдөй	12
23	Кербендерден	12
24	Күйгөндүктөн	12
25	Кыйналдырбай	12
26	Кылчайбастан	12
27	КЫРГЫЗДАРДЫН	12
28	Кырчылдашкан	12
29	Музбурчактын	12
30	Мылтыктуусун	12
31	Сезгенбестен	12
32	Сүйгөндүктөн	12
33	Тайбуурулдай	12
34	Тайбуурулдун	12
35	Тайманбастан	12

**Таблица 5.4.11. Семетей эпосу Жусуп Мамай 2вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	БААТЫРЛАРЫНАН	13
2	Баатырлардын	12
3	Жезкемпирдин	12
4	Кайраттанган	12
5	Калмактардын	12
6	Канжарлуулар	12
7	Карсылдайсың	12
8	Козголбостон	12
9	Көөрүктүүнүн	12
10	Күрпүлдөктөн	12
11	Куткарбастан	12
12	Кызгалдактай	12
13	Кызыталактар	12
14	Кылкылдактын	12
15	Кыргыздардын	12
16	Тайбуурулдун	12
17	Тайманбастан	12
18	ЧЫККЫНЧЫЛЫГЫ	12
19	Айбалтаңдын	11
20	Айдаркандын	11
21	Айталбаймын	11
22	Аккелтеңдин	11
23	Баатырларды	11
24	Баатырларын	11
25	Байтайлакты	11
26	Балталуулар	11
27	Балталуунун	11
28	Барбагандар	11
29	Бастырганда	11
30	Бурчактатып	11
31	Бүтүргөндөй	11
32	Дасторкондо	11
33	Жалбырактай	11
34	Жалманбозду	11
35	Жамгырчынын	11

**Таблица 5.4.12. Семетей эпосу Жусуп Мамай 3вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Таарынттырбай	13
2	Арстандардан	12
3	Арстандыктан	12
4	БААТЫРЛАРДЫН	12
5	Баатырсынган	12
6	Беттешкенден	12
7	Беттештирбей	12
8	Былкылдатпай	12
9	Жазганбастан	12
10	Жалтангансып	12
11	Жезкемпирдин	12
12	Жуукурланган	12
13	Кайсоолойсуң	12
14	Калмактардан	12
15	Калмактардын	12
16	Калтырбастан	12
17	Каңгайлыктын	12
18	карыгандыгым	12
19	Келгендердин	12
20	Кеткендердин	12
21	Козголбостон	12
22	Коркконундан	12
23	Көөрүктүүнүн	12
24	Күйгүлтүктүү	12
25	Күйрөктүүнүн	12
26	Күрпүлдөктүн	12
27	Кыймылдатпай	12
28	Музбурчактын	12
29	Сайдырбастан	12
30	Салтантаттуу	12
31	Сенделтпейли	12
32	Суктанбастан	12
33	Таарынбаймын	12
34	Тайбуурулдай	12
35	Тайбуурулдун	12

**Таблица 5.4.13. Семетей эпосу С.Каралаевдын вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	салоомалейкум	13
2	Темирканындын	13
3	Абийирдүүсүн	12
4	Айткандардын	12
5	Баатырсынган	12
6	Бакдөөлөттөй	12
7	Балбандардын	12
8	Бармактуудан	12
9	Баткактуунун	12
10	Бейдаараттуу	12
11	Беттешкендин	12
12	билгендердин	12
13	Билгичтерден	12
14	Билгичтердин	12
15	Бойбойлошкон	12
16	Буйдалбастан	12
17	Быркылдатпай	12
18	Желдейбоздун	12
19	Желтийбестин	12
20	Жембаштыктай	12
21	жолборсумдан	12
22	Жолборсундун	12
23	Жортормондун	12
24	Жылгындуунун	12
25	Жылмайгырдын	12
26	Кайраттуусун	12
27	Кайшаңдатпай	12
28	Калмактардын	12
29	Канкорлордун	12
30	Каптаптырсың	12
31	караңгысында	12
32	Качкындардын	12
33	Көгөрткөндөн	12
34	Келгендерден	12
35	Коңкулбоздун	12

**Таблица 5.4.14. Сейтек эпосу С.Каралаевдын вариантындагы текст
ичиндеги сөздөрдүн сөз узундугу боюнча жайгашуусу**

№	СӨЗ	УЗУНДУГУ
1	Каарлангандай	13
2	Айткандардай	12
3	Акжолборстой	12
4	Баатырдыктын	12
5	Баатырдырсың	12
6	Баатырлардын	12
7	Баатырсынган	12
8	Байтайлактын	12
9	Бакбурчундун	12
10	Бакчуулардан	12
11	Балбандардай	12
12	Балбандардан	12
13	Балбандардын	12
14	Балбандыктан	12
15	Балбансынган	12
16	Балтыркандын	12
17	Бармактиштеп	12
18	Басташкандын	12
19	Беттешкендин	12
20	Билгендерден	12
21	Билгизбестен	12
22	Билгичтерден	12
23	Болжолсуздун	12
24	Буйдалбастан	12
25	Булкунгандан	12
26	былкылдаарга	12
27	Былкылдатпай	12
28	Бээжиндиктин	12
29	Жадаалданган	12
30	Жадаалдансаң	12
31	Жадаалдашпай	12
32	Жаңжалдашкан	12
33	Жаңжалдашпай	12
34	Жаңжалдашсак	12
35	Желжетпестен	12

5.5. Сөздөрдүн кайталанышы боюнча маалыматтар.

Таблица 5.5.1. Баардык вариант үчүн текст ичиндеги сөздөрдүн кайталануусу

КИТЕП	ВАРИАНТ	СӨЗ	КАЙТАЛАНУУСУ
МАНАС	С.КАРАЛАЕВ	деп	2877
МАНАС	ЖУСУП МАМАЙ	деп	2778
МАНАС	С.КАРАЛАЕВ	экен	2222
МАНАС	С.КАРАЛАЕВ	Манас	2116
МАНАС	С.КАРАЛАЕВ	кан	1761
МАНАС	ЖУСУП МАМАЙ	менен	1566
МАНАС	С.КАРАЛАЕВ	бар	1509
МАНАС	С.КАРАЛАЕВ	эр	1508
МАНАС	С.КАРАЛАЕВ	менен	1488
МАНАС	ЖУСУП МАМАЙ	бар	1457
МАНАС	С.КАРАЛАЕВ	бир	1402
СЕМЕТЕЙ	Б.ЮУНУСАЛИЕВ	деп	1379
МАНАС	С.КАРАЛАЕВ	Кара	1373
МАНАС	С.КАРАЛАЕВ	эми	1307
МАНАС	ЖУСУП МАМАЙ	Бир	1304
МАНАС	С.КАРАЛАЕВ	Кырк	1207
СЕЙТЕК	С.КАРАЛАЕВ	деп	1207
МАНАС	С.КАРАЛАЕВ	болуп	1205
МАНАС	С.КАРАЛАЕВ	жок	1200
МАНАС	Б.ЮУНУСАЛИЕВ 1	деп	1192
МАНАС	С.КАРАЛАЕВ	кытай	1148
МАНАС	ЖУСУП МАМАЙ	экен	1134
МАНАС	Б.ЮУНУСАЛИЕВ 2	деп	1052
СЕМЕТЕЙ	С.КАРАЛАЕВ	деп	1048
МАНАС	С.КАРАЛАЕВ	кылып	991
МАНАС	С.КАРАЛАЕВ	эле	945
МАНАС	ЖУСУП МАМАЙ	эле	928
МАНАС	С.КАРАЛАЕВ	алып	916
МАНАС	С.КАРАЛАЕВ	Көк	874
МАНАС	С.КАРАЛАЕВ	Ал	872
СЕМЕТЕЙ	ЖУСУП МАМАЙ 1	деп	841
МАНАС	ЖУСУП МАМАЙ	Манас	837
МАНАС	Б.ЖАКИЕВ	деп	826
МАНАС	ЖУСУП МАМАЙ	алып	823
МАНАС	ЖУСУП МАМАЙ	Алмамбет	797

**Таблица 5.5.2. Манас эпосу Б.Жакиевдин вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	826
2	Манас	547
3	бар	330
4	менен	324
5	эр	300
6	Бирөө	298
7	экен	291
8	эми	279
9	кан	239
10	Алманбет	229
11	кара	200
12	жок	190
13	алып	185
14	болуп	166
15	кытай	162
16	кырк	154
17	айтат	145
18	Кеп	144
19	кылып	143
20	да	141
21	Эки	140
22	кыргыз	131
23	Жакып	125
24	калган	124
25	кол	123
26	эле	122
27	ат	121
28	калганы	121
29	болду	118
30	баатыр	117
31	бул	114
32	Ал	113
33	келген	112
34	көп	111
35	Алты	108

**Таблица 5.5.3. Манас эпосу Ж.Үмөт уулунун вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	бир	436
2	деп	312
3	менен	235
4	ал	226
5	да	205
6	Хан	169
7	бул	168
8	ошол	129
9	алып	124
10	Мен	117
11	жок	114
12	эле	114
13	деген	111
14	турган	103
15	ар	101
16	бар	100
17	калган	96
18	ушул	96
19	А	92
20	салып	91
21	Сен	89
22	Улуу	89
23	Кырк	87
24	берген	86
25	дагы	84
26	мага	84
27	тең	84
28	үчүн	84
29	күн	80
30	бирок	79
31	Көз	79
32	келген	78
33	кол	78
34	тартып	78
35	экен	78

**Таблица 5.5.4. Манас эпосу Жусуп Мамайдын вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	2778
2	менен	1566
3	бар	1457
4	Бир	1304
5	экен	1134
6	эле	928
7	Манас	837
8	алып	823
9	Алмамбет	797
10	эр	780
11	жок	760
12	ал	595
13	Кара	562
14	болуп	556
15	кеп	548
16	Бул	544
17	Хан	522
18	Эки	486
19	Чубак	482
20	көп	455
21	келип	453
22	айтып	444
23	Мен	441
24	кылып	439
25	ат	426
26	турган	415
27	албай	382
28	келген	379
29	калды	371
30	киши	353
31	болду	352
32	баатыр	349
33	күн	349
34	эл	347
35	деген	342

**Таблица 5.5.5. Манас эпосу С.Каралаевдын вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	2877
2	экен	2222
3	Манас	2116
4	кан	1761
5	бар	1509
6	эр	1508
7	менен	1488
8	бир	1402
9	Кара	1373
10	эми	1307
11	Кырк	1207
12	болуп	1205
13	жок	1200
14	кытай	1148
15	кылып	991
16	эле	945
17	алып	916
18	Көк	874
19	Ал	872
20	көп	764
21	Бул	727
22	Алмамбет	717
23	жер	702
24	эки	675
25	Ошол	666
26	жан	660
27	кол	651
28	Ошондо	634
29	Бакай	603
30	турган	599
31	Калың	585
32	шер	576
33	Чоң	551
34	да	550
35	келген	522

**Таблица 5.5.6. Манас эпосу Б.Юунусалиев 1 вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	1192
2	Манас	577
3	бар	567
4	экен	477
5	менен	472
6	бир	419
7	эми	400
8	жок	360
9	Кара	347
10	эр	329
11	болуп	293
12	Алып	255
13	бул	246
14	Жакып	215
15	келген	208
16	Айтып	200
17	Алмамбет	200
18	Көк	194
19	кеп	186
20	кырк	186
21	кылып	185
22	айтат	182
23	болду	178
24	Көп	176
25	Кан	170
26	жер	169
27	турган	169
28	ал	166
29	болсо	165
30	жан	161
31	эки	160
32	келип	159
33	Баатыр	154
34	Ат	149
35	Кол	149

**Таблица 5.5.7. Манас эпосу Б.Юунусалиев 2 вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	1052
2	бар	561
3	эр	532
4	Кара	473
5	эми	467
6	экен	461
7	Манас	448
8	Бир	431
9	менен	411
10	Алмамбет	352
11	жок	273
12	болуп	257
13	калмак	244
14	көп	243
15	Кан	238
16	Бул	224
17	Коңурбай	222
18	Алып	202
19	көк	198
20	Кошой	196
21	Чубак	193
22	Кырк	183
23	Сыргак	179
24	Ат	178
25	эле	178
26	Ошондо	174
27	турган	174
28	кылып	173
29	Баатыр	168
30	келген	168
31	Ал	164
32	калган	164
33	найза	163
34	калыптыр	159
35	Жер	154

**Таблица 5.5.8. Семетей эпосу Б.Юунусалиевдин вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	1379
2	менен	606
3	жок	508
4	Семетей	429
5	кеп	389
6	Кан	359
7	Кара	356
8	эр	340
9	болуп	339
10	бар	316
11	айтат	309
12	эле	288
13	эми	275
14	калган	264
15	турган	262
16	кылып	251
17	бир	241
18	көп	237
19	бул	219
20	экен	207
21	Алып	206
22	эки	206
23	ошентип	205
24	болсо	192
25	ак	191
26	албай	190
27	дейт	189
28	Каныкей	189
29	Бакай	170
30	эмес	169
31	айтып	168
32	болгон	165
33	калды	164
34	Күлчоро	160
35	келген	155

**Таблица 5.5.9. Семетей эпосу Жусуп Мамай 1 вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	841
2	менен	582
3	Семетей	365
4	бар	341
5	эр	302
6	айтат	284
7	жок	283
8	эми	262
9	алып	253
10	эле	253
11	экен	240
12	Кара	237
13	Хан	232
14	Ал	220
15	кеп	219
16	болуп	218
17	Гүлчоро	205
18	эки	194
19	Манас	190
20	ак	185
21	бир	182
22	Бакай	168
23	ат	166
24	Айчүрөк	158
25	келген	154
26	кылып	146
27	Канчоро	144
28	болгон	142
29	дейт	141
30	Чынкожо	133
31	Айтып	130
32	анда	129
33	Каныкей	125
34	болсо	122
35	бул	122

**Таблица 5.5.10. Семетей эпосу Жусуп Мамай 2 вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	237
2	менен	131
3	бир	86
4	бар	77
5	эле	72
6	Гүлчоро	71
7	жок	71
8	алып	66
9	Канчоро	60
10	Семетей	59
11	эки	58
12	кылып	53
13	Кара	51
14	экен	51
15	Керимбай	50
16	көп	48
17	Эр	47
18	Мен	46
19	уулу	46
20	Ал	45
21	бул	45
22	алды	44
23	хан	43
24	киши	42
25	Бакай	39
26	Кызыл	39
27	чоң	38
28	келип	37
29	баатыр	35
30	кеп	35
31	Эми	35
32	Келген	34
33	Карагул	33
34	күн	33
35	Ат	31

**Таблица 5.5.11. Семетей эпосу Жусуп Мамай 3 вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	624
2	менен	327
3	Семетей	232
4	эле	225
5	бир	206
6	жок	200
7	Гүлчоро	198
8	Хан	195
9	Канчоро	191
10	Эр	186
11	алып	183
12	Ал	176
13	мен	155
14	экен	154
15	Кара	149
16	Сен	135
17	бул	127
18	баатыр	125
19	бар	124
20	Жалгыз	124
21	кылып	120
22	болуп	116
23	Кеп	109
24	Мадыкан	102
25	калган	100
26	келген	94
27	салып	94
28	албай	93
29	киши	93
30	Эки	93
31	Кызыл	92
32	тартып	89
33	Эми	89
34	Көп	88
35	келип	86

**Таблица 5.5.12. Семетей эпосу С.Каралаевдын вариантындагы текст
ичиндеги сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	1048
2	Кан	675
3	Бир	636
4	эле	615
5	эми	553
6	Бакай	537
7	менен	528
8	жок	521
9	калган	501
10	кара	482
11	бар	472
12	Каныкей	462
13	экен	458
14	Манас	447
15	кырк	420
16	Семетей	371
17	кылып	366
18	ал	362
19	болуп	362
20	Эр	334
21	алып	297
22	кайран	295
23	Мен	288
24	Алты	285
25	турган	254
26	жер	248
27	Ошондо	248
28	Атаң	246
29	эки	246
30	албай	243
31	Ат	237
32	жан	233
33	жетим	230
34	Ак	223
35	Кудай	216

**Таблица 5.5.13. Сейтек эпосу С.Каралаевдын вариантындагы текст ичиндеги
сөздөрдүн кайталануусу**

№	СӨЗ	КАЙТАЛАНУУСУ
1	деп	1207
2	эр	750
3	болуп	725
4	экен	601
5	эле	577
6	кылып	493
7	кара	481
8	менен	461
9	бир	443
10	бар	378
11	кан	377
12	жер	369
13	Ал	364
14	жок	364
15	жан	358
16	алып	329
17	Сарыбай	317
18	Сейтек	300
19	эки	261
20	Кыяз	257
21	Күлчоро	255
22	Эл	254
23	Кайран	236
24	Ак	232
25	Ат	226
26	Айза	223
27	салып	219
28	уулу	218
29	шер	218
30	кол	216
31	көп	208
32	Чын	208
33	болгон	206
34	калган	201
35	Бу	200

5.6. Программа үчүн берилиштер базасын түзүү.

Программа үчүн MS SQL берилиштер базасын тандап алдым. Анткени бул колдонууга ыңгайлуу жана жасалып жаткан программанын талаптарына жооп берет.

5.7. Программалык кодду түзүү.

Программа веб тармагында иштөө үчүн жасалган. Ошондуктан PHP/JS/JQuery программаларын тандап алдым. Саптагы сөздү таап алуу үчүн төмөнкү функция жасалды.

```
function poisk_podstroki_v_stroke($stroka, $podstroka)  
{  
    $m = false;  
    $str = "<td align='left'>";  
    $massiv = explode(" ", $stroka);  
    $sani = count ($massiv);  
    for($i = 0; $i < $sani; $i++){  
        if($massiv[$i] == $podstroka){  
            $m = true;  
            $str = $str."</td><td  
align='left'><b>".$massiv[$i]."</b></td><td align='left'>";  
        }  
        else  
        {  
            $str = $str.$massiv[$i]." ";  
        }  
    }  
}
```

```

    }
    if ($m)
        return $str."</td>";
    else
        return 1;
}

```

Жогорудагы функцияда берилген сапты explode функциясы аркылуу бир массивге топтойт жана ал жерден izdeluuchy сөздү издеп баштайт. Эгерде изделген сөз бар болсо ошол сапты толугу менен чыгарып берет.

Диаграммалар үчүн pChart библиотекасы колдонулду жана анын жардамы менен төмөндөгүдөй коддор жазылды.

```

$DataSet = new pData();

$Requete = "SELECT Kitep, Variant, COUNT(texti) AS san,
COUNT(DISTINCT texti) AS sandistinct FROM Sozdor_new GROUP BY
Kitep, Variant ORDER BY Kitep, Variant";

$result = mssql_query($Requete,$connect);

while($row = mssql_fetch_array($result))
{
    $DataSet->AddPoint($row["Variant"],"Variant");

    $DataSet->AddPoint(round ($row["san"]/$row["sandistinct"],2),"ort");
}

$DataSet->AddAllSeries();

```

```

$DataSet->SetAbsciseLabelSerie("Variant");

// Initialise the graph

$Test = new pChart(680,200);

$Test->drawFilledRoundedRectangle(7,7,473,193,5,240,240,240);

$Test->drawRoundedRectangle(5,5,475,195,5,230,230,230);

// Draw the pie chart

$Test->setFontProperties("Fonts/tahoma.ttf",8);

$Test->drawPieGraph($DataSet->GetData(),$DataSet-
>GetDataDescription(),150,90,110,PIE_PERCENTAGE,TRUE,50,20,5);

$Test->drawPieLegend(310,15,$DataSet->GetData(),$DataSet-
>GetDataDescription(),250,250,250);

$Test->Render("Genel.png");

```

5.8. Колдонуучуга маалымат.



Сүрөт2. Конкорданс программасынын иштөө чөйрөсү(Манас эпосунун варианттарына)

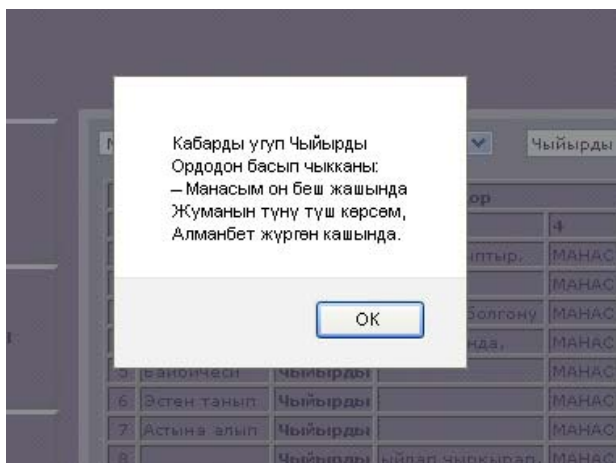
Жогорудагы сүрөттө веб баракчасына кирүү бети көрсөтүлгөн. Бул жерге колдонуучулар жана башкаруучу(администратор) кире

алат. Башкаруучу китептерди киргизүү мүмкүнчүлүгүнө ээ. Башкы бет аркылуу кирилгенден кийин китепти, вариантты тандап, изделүүчү сөздү жазуу керек.

#	1	2	3	4	5
1		Чыйырды	атап калыптыр.	МАНАС	Б.ЖАКИЕВ
2	Байбичеси	Чыйырды		МАНАС	Б.ЖАКИЕВ
3		Чыйырды	кош бой болгону	МАНАС	Б.ЖАКИЕВ
4		Чыйырды	аны укканда,	МАНАС	Б.ЖАКИЕВ
5	Байбичеси	Чыйырды		МАНАС	Б.ЖАКИЕВ
6	Эстен танып	Чыйырды		МАНАС	Б.ЖАКИЕВ
7	Астына алып	Чыйырды		МАНАС	Б.ЖАКИЕВ
8		Чыйырды	ыйлап чыкырап,	МАНАС	Б.ЖАКИЕВ
9	- О,	Чыйырды	менен Жакыбым,	МАНАС	Б.ЖАКИЕВ
10		Чыйырды	көзүн жаштады.	МАНАС	Б.ЖАКИЕВ
11	Кабарды угуп	Чыйырды		МАНАС	Б.ЖАКИЕВ
12		Чыйырды	эне баш болуп,	МАНАС	Б.ЖАКИЕВ
13	Кудагыый	Чыйырды		МАНАС	Б.ЖАКИЕВ
14	Энекеси	Чыйырды		МАНАС	Б.ЖАКИЕВ
15	Ойрон энең	Чыйырды		МАНАС	Б.ЖАКИЕВ

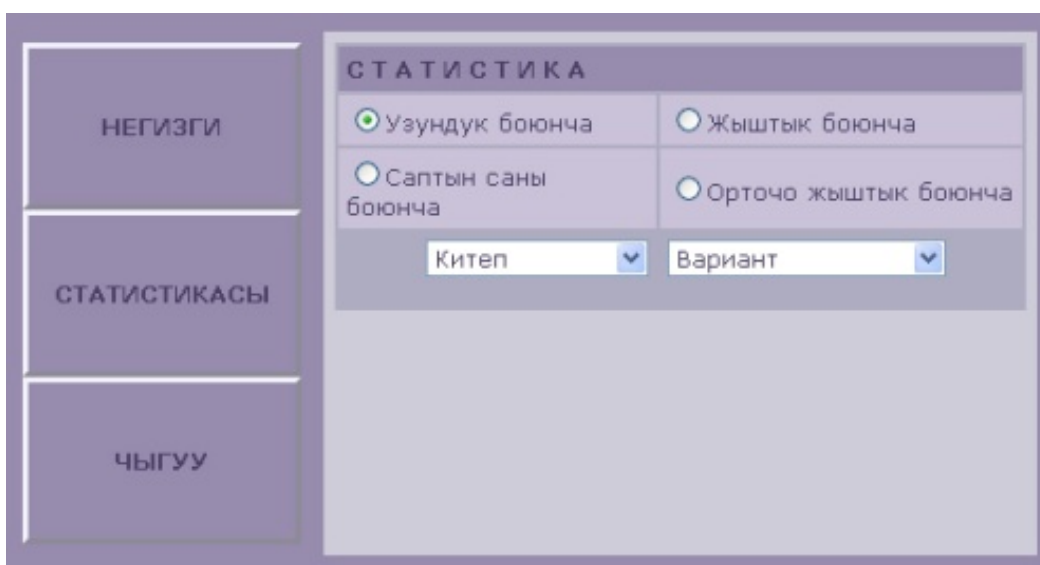
Сүрөт 3. Эпостогу сөздөрдү издөө

Жогоруда изделген сөз боюнча тизме көрсөтүлдү. Бул тизме изделген сөз катышкан сапты гана көрсөтөт. Бул тизменен ар бир сабына басканыбызда ошол сап катышкан сүйлөм толугу менен көрсөтүлөт. Ал төмөндөгү сүрөттө көрсөтүлгөн.



Сүрөт 4. Сөздөр боюнча сүйлөмдүн экранга чыгуусу

Статистикалык маалыматтар бөлүмүндө Манас эпосуна байланышкан бир нече статистика таблица жана графиктер менен көрсөтүлгөн.



Сүрөт 5. Статистиканын негизги бети

НЕГИЗГИ

СТАТИСТИКАСЫ

ЧЫГУУ

СТАТИСТИКА

☒ Узундук боюнча
 ☐ Жыштык боюнча

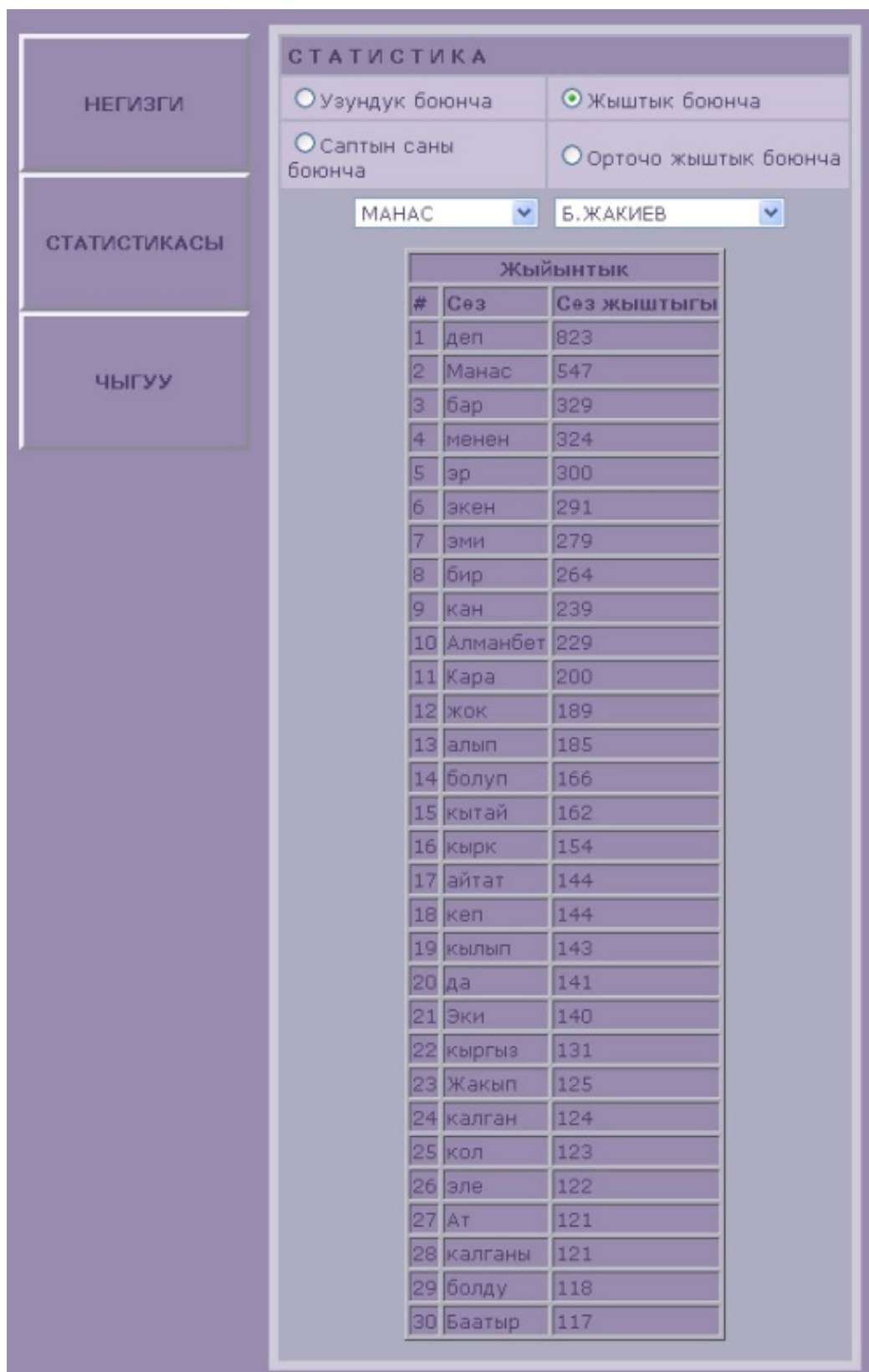
☐ Саптын саны боюнча
 ☐ Орточо жыштык боюнча

МАНАС

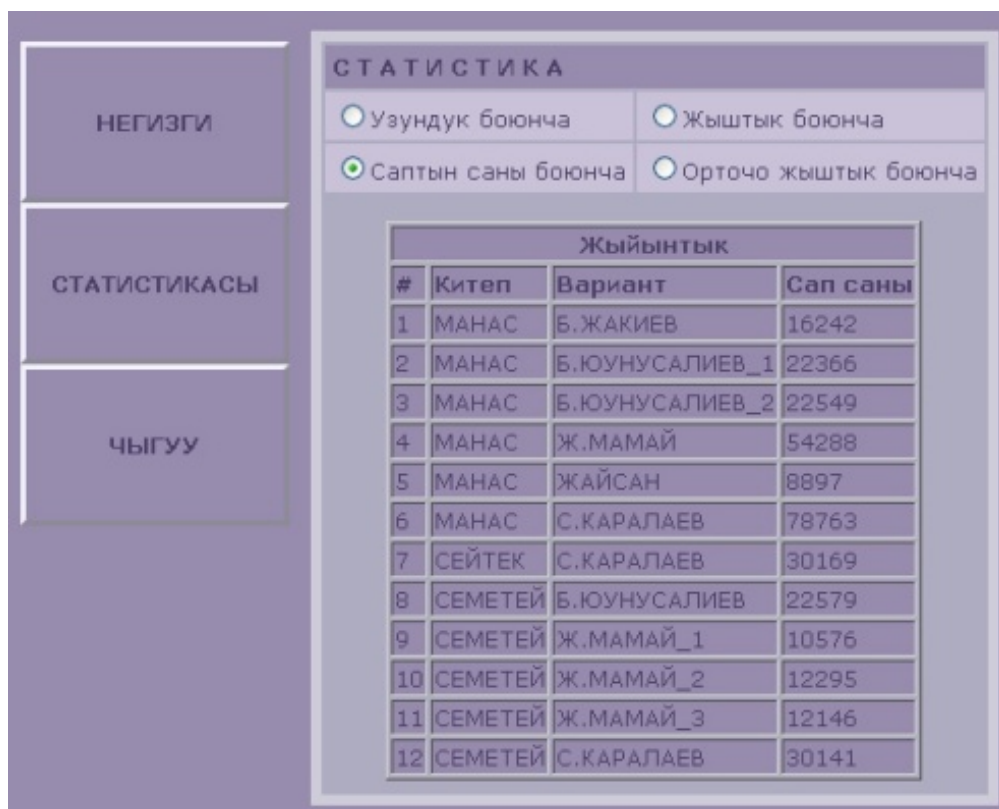
Б.ЖАКИЕВ

Жыйынтык		
#	Сөз	Сөз узундугу
1	Кабыландардан	13
2	Кабыландардын	13
3	пейилчечилдик	13
4	тердеткенинде	13
5	Айттырбастан	12
6	Арстандардын	12
7	БААТЫРЛАРДЫН	12
8	Баатырсынган	12
9	Бакдөөлөттүн	12
10	Билгендерден	12
11	Билдирбестен	12
12	Биртуугансың	12
13	Бозкертиктин	12
14	Буудайгүлдөн	12
15	Бээжиндиктин	12
16	Жайсаңбайдын	12
17	Жалгыздыктан	12
18	Жергенттикке	12
19	Жеткиргейсиң	12
20	Жолборсуңдун	12
21	Жолдоштордун	12
22	Жортуулчудай	12
23	Жумшаргансып	12
24	Жылгыңдуунун	12
25	Көкжалдардын	12
26	Көрсөтпөймүн	12
27	көтөрүлбөгөн	12
28	Күчтүүсүнбөй	12
29	Күчтүүсүнгөн	12
30	Каңгайлардан	12

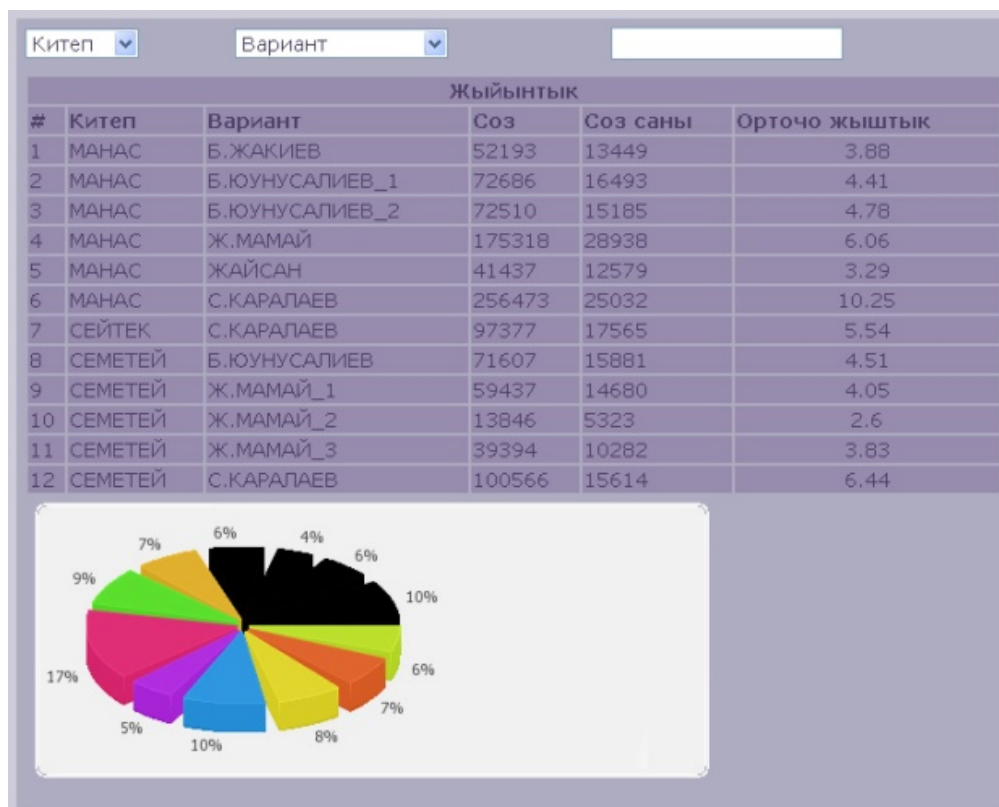
Сүрөт 6. Сөз узундугу боюнча статистика



Сүрөт 7. Сөз жыштыгы боюнча статистика



Сүрөт 8. Саптын саны боюнча статистика



Сүрөт 9. Орточо жыштык боюнча статистика

КОРУТУНДУ.

Корпустук лингвистика лингвистика билиминин жаңы багыты болуп тилдин ар бирдигинин кайсыл учурда кандай мааниде колдонулгандыгын изилдегенге мүмкүнчүлүк берет. Улуттук жана атайын багытта түзүлгөн корпустар изилдөөчүлөргө автоматтык түрдө издөө, эмпирикалык материалдарды системалоо, чоң массивдүү маалыматтарды атайын компьютердик жабдыктардын жардамы менен иштеп чыгууга мүмкүнчүлүк түзөт. Корпустук лингвистиканы карасак – бул лингвистикалык изилдөө методологиясы. Бул тез өнүгүп келе жаткан тармак, компьютердик технологияларынын прогресси менен тексттерди автоматтык түрдө иштеп чыгуу жабдык-программаларын түзүүсү менен бирге келет.

Манас эпосу көптөгөн тармактарда изилденип келүүдө. Бирок корпустук лингвистикада изилдөөлөр саналуу гана. Ошондуктан магистрдик ишимди Манас эпосунун корпустарынын негизинде иштөөнү туура таптым.

Бара бара учурдагы программалар четкы чыгып алардын ордуна дагы да жакшы, күчтүү ийкемдүү, көп функционалдуу жаны улуттук корпустар келээри азыр деле ишенимдүү.

Колдонулган булактар

- [1]. Б.М.Юнусалиев. Манас. Биринчи бөлүк. I китеп. Фрунзе, Кыргызмамбас, 1958.
- [2]. Б.М.Юнусалиев, Манас. Биринчи бөлүк. II китеп. Фрунзе, Кыргызмамбас, 1958.
- [3]. Б.М.Юнусалиев, Семетей Манас эпосунун экинчи бөлүгү. III китеп. Фрунзе, Кыргызмамбас, 1959.
- [4]. Б.М.Юнусалиев, Сейтек Манас эпосунун үчүнчү бөлүгү. IV китеп. Фрунзе, Кыргызмамбас, 1960.
- [5]. Баранов А.Н. Введение в прикладную лингвистику: – М.: ЛКИ, 2007.
- [6]. Б. Жакиев Манас кыргыздардын баатырдык эпосу. Бийиктик, 2011.
- [7]. Ж.Үмөт уулу Айкөл манасБиринчи китеп. Бийиктик, 2005.
- [8]. Садыков Т., Шаршембаев Б. Манас; Кыргызча-Түркчө чоң көрсөткүч сөздүк. Анкара,2011.-1647 б.
- [9]. Саякбай Каралаев Манас. Бишкек, Турар, 2010.
- [10].Саякбай Каралаев Семетей. Фрунзе, Кыргызполиграфкомбинат, 1987.
- [11].Саякбай Каралаев Сейтек. Бишкек,
- [12]. Гвишиани Н.Б. Практикум по корпусной лингвистике: Учеб. пособие по английскому языку. – М.: Издательство Высшая школа, 2008.
- [13]. Захаров В.П. Корпусная лингвистика. Учебно-метод. пособие. – СПб.: Издательство СПбГУ, 2006.
- [14]. Захаров В.П. Информационно-поисковые системы: Учебно-метод. пособие. — СПб.: Издательство СПбГУ, 2005.
- [15]. Рыков В.В. Курс лекций по корпусной лингвистике. [Электронный ресурс]. – Режим доступа: <http://rykov-cl.narod.ru/c.html>.
- [16]. Баранов А.Н. Введение в прикладную лингвистику: Учеб. пособие. – М.: ЛКИ, 2007.
- [17]. Белозёрова Н.Н., Чуфистова Л.Е. Шекспир и компания, или использование электронных библиотек при лингвистическом исследовании: Учеб. пособие. – Тюмень: Издательство ТюмГУ, 2007.

- [18]. Зубов А.В. Информационные технологии в лингвистике: Учеб. пособие. – М.: Издательский центр «Академия», 2004.
- [19]. Информационные системы (документальный поиск): Учебное пособие. – СПб.: Издательство СПбГУ, 2002.
- [20]. Информационные технологии: Основы работы с программными средствами и информационными ресурсами Интернет: Практикум / Сост. Захаров В.П., Арбатская О.А.; Под ред. В.П. Захарова; Санкт-Петербург. Гос. ун-т культуры и искусств. – СПб.: Издательство СПбГУКИ, 2003.
- [21]. Коваль С.А. Лингвистические проблемы компьютерной морфологии. – СПб.: Издательство С.-Петерб. ун-та, 2005.
- [22]. Марчук Ю.Н. Основы компьютерной лингвистики. – М.: Издательство МПУ, 2000.
- [23]. Корпуса, созданные при поддержке Программы фундаментальных исследований Президиума РАН «Корпусная лингвистика». <http://web-corpora.net>, Erişim tarihi: 15Nisan 2015.
- [24]. Национальный корпус русского языка. <http://www.ruscorpora.ru>, Erişim tarihi: 17Nisan 2015.

ӨМҮР БАЯН

Аты жөнү:

Уйчубек кызы Мээрим

Туулган жери жана жылы:

Кыргыз Республикасы, Ысыккөл облусу, Тоң району, Караталаа айылы 1981- ж. 24-февраль

Окуган мекемелер:

Мектеп: В.И.Ленин атындагы орто мектеп

Бакалавриат: Кыргыз-Түрк «Манас» Университети, Инженерия факультети, Компьютер инженерия бөлүмү

Бакалавриат: Ж.Баласагын атындагы Кыргыз улуттук университет, Экономика жана тармактык башкаруу факультети, Экономика бөлүмү

Магистратура: Кыргыз-Түрк «Манас» Университети, Табигый Илимдер Институту, Компьютер инженерия билим тармагы

Иштеген мекемелер:

1. “ТАШГАР” *Жоопкерчилиги чектелген коому, ПК Оператору*
2. Кыргыз-Түрк «Манас» Университети, Студенттик бөлүмү, Программист

Билген тилдери жана анын деңгээли:

Кыргызча(эне тил, эң жакшы), орусча(жакшы), түркчө(жакшы), англисче(жакшы)

Үй-бүлөлүк абалы: Үй-бүлөлүү, Бир баласы бар

Кол: _____