

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI  
VA ADABIYOTI UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR  
BERUVCHI DSc.03/30.12.2019.Fil.19.01 RAQAMLI ILMIY KENGASH  
ASOSIDAGI BIR MARTALIK ILMIY KENGASH**

---

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI  
VA ADABIYOTI UNIVERSITETI**

**YODGOROV UMIDJON SAYDILLA O‘G‘LI**

**O‘ZBEK TILI KORPUSIDA FRAZEMALARNI TEGGLASHNING  
LINGVISTIK VA DASTURIY ASOSLARI**

**10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi**

**Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi**

**AVTOREFERATI**

**Toshkent – 2025**

**Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi  
avtoreferati mundarijasi**

**Contents of dissertation of sciences of the doctor of philosophy (PhD) on  
philological sciences**

**Оглавление автореферата диссертации доктора философии (PhD) по  
филологическим наукам**

**Yodgorov Umidjon Saydilla o‘g‘li**

O‘zbek tili korpusida frazemalarni teglashning lingvistik va dasturiy  
asoslari ..... 3

**Ёдгоров Умиджон Сайдилла ўғли**

Лингвопрограммные основы фразового тегирования в корпусе  
узбекского языка..... 29

**Yodgorov Umidjon Saydilla ugli**

Linguistic support and software for phrasal tagging in the Uzbek  
language corpus ..... 58

**E‘lon qilingan ishlar ro‘uxati**

List of published works  
Список опубликованных работ..... 62

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI  
VA ADABIYOTI UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR  
BERUVCHI DSc.03/30.12.2019.Fil.19.01 RAQAMLI ILMIY KENGASH  
ASOSIDAGI BIR MARTALIK ILMIY KENGASH**

---

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI  
VA ADABIYOTI UNIVERSITETI**

**YODGOROV UMIDJON SAYDILLA O‘G‘LI**

**O‘ZBEK TILI KORPUSIDA FRAZEMALARNI TEGGLASHNING  
LINGVISTIK VA DASTURIY ASOSLARI**

**10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi**

**Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi**

**AVTOREFERATI**

**Toshkent – 2025**

**Falsafa doktori (PhD) dissertatsiyasi mavzusi O‘zbekiston Respublikasi Oliy attestatsiya komissiyasida B2023.3.PhD/Fil3963 raqam bilan ro‘yxatga olingan.**

Dissertatsiya Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o‘zbek, ingliz, rus (rezyume) Ilmiy kengashning veb-sahifasi ([www.tsuull.uz](http://www.tsuull.uz)) va “Ziyonet” Axborot ta’lim portalida ([www.ziyonet.uz](http://www.ziyonet.uz)) joylashtirilgan.

**Ilmiy rahbar:**

**Elov Botir Boltayevich**

texnika fanlari bo‘yicha falsafa doktori, dotsent

**Rasmiy opponentlar:**

**Abdurahmonova Nilufar Zaynobiddin qizi**

filologiya fanlari doktori (DSc), professor

**Eshmuminov Asqar Allamurodovich**

filologiya fanlari doktori (DSc), dotsent

**Yetakchi tashkilot:**

**Samarqand davlat universiteti**

Dissertatsiya himoyasi Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti huzuridagi DSc.03/30.12.2019.Fil.19.01 raqamli Ilmiy kengash asosidagi bir martalik Ilmiy kengashning 2024-yil “\_\_\_” \_\_\_\_\_ soat \_\_\_\_\_dagi majlisida bo‘lib o‘tadi. (Manzil: 100100, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko‘chasi, 103. Tel: (99871) 281-42-44; faks: (99871) 281-42-44; faks: (99871) 281-12-44. ([www.tsuull.uz](http://www.tsuull.uz)). e-mail: [monitoring@navoiy-uni.uz](mailto:monitoring@navoiy-uni.uz)).

Dissertatsiya bilan Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti Axborot-resurs markazida tanishish mumkin. ( -raqam bilan ro‘yxatga olingan) (Manzil: 100100, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko‘chasi, 103. Tel: (99871) 281-42-44; faks: (99871) 281-42-44; faks: (99871) 281-12-44 ([www.tsuull.uz](http://www.tsuull.uz)).

Dissertatsiya avtoreferati 2025-yil “\_\_\_” \_\_\_\_\_kuni tarqatildi.

(2025-yil “\_\_\_” \_\_\_\_\_dagi \_\_\_\_\_ raqamli reestr bayonnomasi)

**H. Dadaboyev**

Ilmiy darajalar beruvchi ilmiy kengash asosidagi bir martalik ilmiy kengash raisi, filol.f.d., professor

**Q.U. Pardayev**

Ilmiy darajalar beruvchi ilmiy kengash asosidagi bir martalik ilmiy kengash ilmiy kotibi, filol.f.d., professor

**S.BE. Normamatov**

Ilmiy darajalar beruvchi ilmiy kengash asosidagi bir martalik ilmiy kengash qoshidagi ilmiy seminar raisi, filol.f.d., professor

## **KIRISH (Falsafa doktori (PhD) dissertatsiyasi annotatsiyasi)**

**Tadqiqot mavzusining dolzarbligi va zarurati.** Jahon kompyuter lingvistikasida tabiiy tilga avtomatik ishlov berishda turli til birliklari, jumladan, leksik birliklarni lingvistik teglash muammosi tabiiy tilga ishlov berish, matematik lingvistika va mashina tarjimasida amaliyotida faol qo'llanmoqda. Turli tillarning teglash dasturini ishlab chiqishda frazeologik birliklarni teglash masalasi alohida muammo sifatida qaraladi. Fandagi bu yangilanishlar axborot texnologiyalarini tilshunoslikka tatbiq etish bilan bog'liq istiqbolli ilmiy yo'nalishlarning vujudga kelishiga asos bo'ldi. Zero, bu axborot-qidiruv tizimi, korpus lingvistikasi va mashina tarjimasida morfologik tahlilning umumiy tamoyillarini o'rganish, tahlil tizimlari lingvistik ta'minoti uchun formal til va grammatikani yaratish tizimini ishlab chiqishning zaruriyatini yuzaga keltirdi. Bu esa korpusda frazeologik birliklarni qidirish, avtomatik tarjimada frazeologik birliklarni yaxlit birlik sifatida tanib olish, tezauruslarni ishlab chiqish borasidagi imkoniyatlarni kengaytiradi.

Dunyo tilshunosligida XX asr oxirida tilni lingvistik modellashtirish, til korpuslarida lisoniy birliklarni avtomatik tanib olish, muayyan tillar doirasida frazeologik birliklarni lemmalash, teglash nazariyasi va amaliyoti, algoritmini yaratish kabi masalalar kun tartibiga qo'yildi. Frazeologizmlarni avtomatik teglash, korpusda ularni ajratib ko'rsatish tahlilni soddalashtiradi. Frazemalarni lemmatizatsiya qilish ularning korpuslarda qo'llanilishini statistik tahlil qilishni osonlashtiradi, bu frazemalarning semantikasi, pragmatikasi va uslubiy xususiyatlari haqida asosli xulosalar chiqarishga imkon beradi. Mashina tarjimasida sentiment tahlil va ma'lumotlarni ajratib olish kabi tabiiy tilga avtomatik ishlov berish tizimlarini yaratish yo'lidagi muhim qadamdir. Til korpusida frazeologik birliklarning ajratib ko'rsatilishi (teglanishi) frazeologiyani o'rganish, jumladan, frazemalarning qo'llanilish qonuniyatlarini, ularning semantik va strukturaviy xususiyatlarini aniqlashga yordam beradi. Lemmatizatsiya qilingan frazemalar korpuslari frazeologik lug'atlar, o'quv qo'llanmalari va til o'rganish uchun dasturiy ta'minot kabi turli xil lingvistik resurslarni yaratishda ishlatiladi.

O'zbek kompyuter lingvistikasi sohasida til korpusi, avtomatik tarjima, sun'iy intellekt, o'zbek tilini tushunish va qayta ishlash borasida ilmiy izlanishlarga e'tibor qaratilmoqda. Lekin o'zbek tili korpusida frazeologik birliklarni avtomatik teglash algoritmi va modelini ishlab chiqish masalasi ilmiy-nazariy planda o'rganilmagan. Zero, "...davlat tilining sofligini saqlash, uni boyitib borish va aholining nutq madaniyatini oshirish; davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta'minlash"<sup>1</sup> hozirgi kunda o'zbek kompyuter lingvistikasi oldida turgan dolzarb vazifalardan biridir. Shunga ko'ra, o'zbek tiliga zamonaviy axborot texnologiyalari vositasida ishlov berilishiga erishish, buning uchun til korpuslari, elektron tarjimon, tezaurus kabi dastlabki avtomatik ishlov vositalari hamda ularning lingvistik ta'minotini yaratish singari masalalar muhim ahamiyat kasb etadi.

---

<sup>1</sup> O'zbekiston Respublikasi Prezidentining 2019-yil 21-oktyabrdagi "O'zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to'g'risida"gi PF-5850-son farmoni. <https://lex.uz/docs/-4561730>

O‘zbekiston Respublikasi Prezidentining 2016-yil 13-maydagi “Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetini tashkil etish to‘g‘risida”gi PF-4997-son, 2017-yil 17-fevraldagi “Fanlar akademiyasi faoliyati, ilmiy-tadqiqot ishlarini tashkil etish, boshqarish va moliyalashtirishni yanada takomillashtirish chora-tadbirlari to‘g‘risida”gi PF-2789-son, 2019-yil 21-oktyabrdagi “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to‘g‘risida”gi PF-5850-son, 2020-yil 20-oktyabrdagi “Mamlakatimizda o‘zbek tilini yanada rivolantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi PF-6084-son farmonlari, 2019-yil 4-oktyabrdagi “O‘zbekiston Respublikasining “Davlat tili haqida”gi Qonuni qabul qilinganligining o‘ttiz yilligini keng nishonlash to‘g‘risida”gi PQ-4479-son qarori, 2020-yil 24-yanvardagi O‘zbekiston Respublikasi Prezidentining Oliy Majlisga Murojaatnomasi hamda boshqa me‘yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu tadqiqot muayyan darajada xizmat qiladi.

**Tadqiqotning respublika fan va texnologiyalari taraqqiyoti ustuvor yo‘nalishlariga mosligi.** Tadqiqot respublika fan va texnologiyalari rivojlanishining I. “Axborotlashgan jamiyat va demokratik davlatni ijtimoiy, huquqiy, iqtisodiy, madaniy, ma‘naviy-ma‘rifiy rivojlantirishda innovatsion g‘oyalar tizimini shakllantirish va ularni amalga oshirish yo‘llari” ustuvor yo‘nalishiga muvofiq bajarilgan.

**Muammoning o‘rganilganlik darajasi.** Jahon korpus lingvistikasida til korpusini tuzish, korpus birliklarini avtomatik lingvistik teglashga oid masalalar G.I.Kustova, O.N.Lyashevskaya, Y.V.Paducheva, Y.V.Raxilinalarning tadqiqotlarida tahlil qilingan. Frazeologik birliklarning lisoniy xususiyatlari tadqiqiga bag‘ishlangan ishlar sirasida A.Ketrin<sup>2</sup>, Y.Degsdottir<sup>3</sup>, R.Sarif<sup>4</sup>ning ishlari e‘tiborga loyiq. Shuningdek, nutqda idiomatik birliklarning qo‘llanishi<sup>5</sup>, ingliz va chex tillari frazeologik birliklarning tipologik xususiyatlari<sup>6</sup>, Amerika ingliz tili frazemalarining korpusga asoslangan tadqiqi<sup>7</sup> monografik planda o‘rganilgan.

O‘zbek tilshunosligida frazeologik birliklarning til xususiyati va pragmatik tadqiqiga bag‘ishlangan qator tadqiqotlar mavjud. I.Qo‘chqortoyev, M.Ro‘ziqulova, K.Hayitmetov, B.Yo‘ldoshevlarning tadqiqotlari diqqatga sazovor<sup>8</sup>.

---

<sup>2</sup> Kathryn A. A Study of Idioms in Relation to Language Universals. A Senior Thesis submitted in partial fulfillment of the requirements for graduation in the Honors Program. – Liberty University: Fall, 2015. – Pages 47.

<sup>3</sup> Ylva Dagsdóttir. The teaching of idiomatic expressions in upper secondary school after the implementation of the LK20, from the perspectives of English language teachers. Master’s Thesis. – Lektorutdanning i språkfabrik, 2022. – Pages 92.

<sup>4</sup> Rifka Nur‘aini Syarif. An analysis of idiomatic expression meaning in frozen film script and its implication in teaching vocabulary. State institute for islamic studies of metro, 2020. – Pages 127.

<sup>5</sup> Elizabeth Jane Holt. Figures of speech: an exploration of the use of idiomatic phrases in conversation. – The University of York, Department of Sociology. March, 1991. – Pages 291.

<sup>6</sup> Bc. Ondřej Čada A Comparative Study of English and Czech Idioms Related to Travel, Transport and Motion: Lexical and Syntactical Means. MASARYK UNIVERSITY BRNO, Brno 2018.

<sup>7</sup> Rafatbakhsh, E., Ahmadi, A. A thematic corpus-based study of idioms in the Corpus of Contemporary American English. Asian. J. Second. Foreign. Lang. Educ. 4, 11 (2019). <https://doi.org/10.1186/s40862-019-0076-4>

<sup>8</sup> Кучкартаев И. Фразеологическое новаторство Абдулла Каххара: Автореф. дис. канд. филол. наук. – Ташкент, 1965. 141 б.; Рузикулова М. Идиоматика узбекского языка: автореф. дисс. канд. филол. наук. – Самарканд, 1966, – С. 22; Хайитметов К. Фразеологик бирликларда сўзлар тартиби масаласига доир // ЎТА, 1966. – №2. – Б.45-53.; Каххарова Х. Фразеология Абдуллы Кадыри: автореф. дисс. канд. филол. наук. – Ташкент, 1985, – 24

Sh.Rahmatullayev hozirgi o'zbek tili fe'l iboralarining asosiy grammatik xususiyatlari va o'zbek tilida fe'l frazemalarining bog'lashuvi<sup>9</sup>, A.Mamatov o'zbek tili frazeologizmlarining shakllanishi va frazeologik stilistikasi<sup>10</sup>, A.Mamatov, D.Ochilov frazemalarning sintaktik va derivatsion modellari<sup>11</sup>, G.Jumanazarova frazemalarning nutqda qo'llanilishi<sup>12</sup>, B.Yo'ldoshev hozirgi o'zbek adabiy tilida frazeologik birliklarning funksional-uslubiy xususiyatlari, M.Yuldoshev Cho'lponning xalq iboralaridan foydalanish mahorati<sup>13</sup> masalalarini tadqiq etishgan<sup>14</sup>. Sh.Almatova, M.Vafoyeva, A.Nasirov, Sh.Ganiyevalar frazeologik birliklarni turli aspekt<sup>15</sup>da tadqiq etganini ta'kidlash joiz.

O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari Sh.Xamroyeva, o'zbek tili milliy korpusining sinonim so'zlar bazasi A.Eshmo'minov, o'zbek tili frazeologik birliklarini til korpusida aks ettirish va lingvistik ma'lumotlar bazasini shakllantirish muammosi G.Begmatova, o'zbek tili korpusini tuzish tamoyillari N.Abduraxmonova, G.Toirova, o'zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari O.Abdullayeva, o'zbek tili ta'limiy korpusini tuzishning lingvistik asoslari O'.Xoliyorov, o'zbek tili Milliy korpusini yaratishda kompyuter usullari A.Raxmanova, o'zbek tili korpusi birliklarining uslubiy teglarini yaratish tamoyillari va lingvistik ta'minoti D.Elova, Alisher Navoiy mualliflik korpusi va uning semantik teglari bazasini yaratish N.G'ulomova, o'zbek tili korpusi uchun zamonni ifodalovchi grammatik shakllarni annotatsiyalash E.Xonnazarov, o'zbek tili birliklarini birliklarini tokenlash, stemlash, lemmalash masalasi Z.Xusainova, atov birliklarini o'zbek tili korpusi uchun leksik-semantik

---

с; Йўлдошев Б. Ҳозирги ўзбек адабий тилида фразеологик бирликларнинг функционал-услубий хусусиятлари: филол. фанлари д-ри дисс. автореф. – Тошкент, 1993. – 50 б.; Усмонова Ш. Фразеологик бирликларда маъно ва курилиш номуносивлиги // ЎТА, 1998. – №1. – Б.77.

<sup>9</sup> Раҳматуллаев Ш. Некоторые вопросы узбекской фразеологии: Автореф. ди сс.... д-ра. филол. наук. – Ташкент, 1966. – 47 с.; Раҳматуллаев Ш. Основные грамматические особенности образных глагольных фразеологических единиц современного узбекского языка: Автореф. дисс... канд. филол. наук. – М., 1952. – 19 с.; Раҳматуллаев Ш. Ўзбек тилида феъл фраземаларининг боғлашуви. – Тошкент: Университет, 1992. – 126 б.

<sup>10</sup> Маматов А.Э. Ўзбек тили фразеологизмларининг шаклланиши масалалари: филология фанлари доктори диссертацияси автореферати. – Тошкент, 2000. – 254 б.; Маматов А. Фразеологик стилистика масалари. – Тошкент, 1991. – 151 б.; Маматов А. Фразеологизмларнинг шаклланиш асослари. – Тошкент: Навруз, 1996. – 121 б.

<sup>11</sup> Маматов А., Очилов Д. Фраземаларнинг синтактик ва деривацион моделлари // Стилистика ва прагматика, 2-кисм. – Самарканд: СамДУ, 2010. – Б. 56-57.

<sup>12</sup> Жуманазарова, Г. Фраземаларнинг нутқда қўлланилиши / Г.Жуманазарова [ва бошқ]. – Тошкент: Турон zamin ziyo. 2017. – 128 б.

<sup>13</sup> Йўлдошев М. Чўлпоннинг халқ ибораларидан фойдаланиш маҳорати // Ўзбек тили ва адабиёти. 1999. 6-сон. 12-13-б.

<sup>14</sup> Йўлдошев Б. Ҳозирги ўзбек адабий тилида фразеологик бирликларнинг функционал-услубий хусусиятлари. Филол. фан. д-ри. дисс. – Тошкент, 1993.

<sup>15</sup> Алматова Ш. Ўзбек тили фраземаларининг компонент таҳлили: филол. фан. ном. ... дис. автореф. – Тошкент: ЎЗР ФА ТАИ, 2008. – 22 б.; Вафоева М. Ўзбек тилида фразеологик синонимлар ва уларнинг структурал-семантик таҳлили: филолог. фан. номз. диссер. автореф. – Тошкент, 2009. – 35 б.; Насиров А. Француз, ўзбек ва рус тилларидаги проverbs фразеологизмларнинг семантик-стилистик ва миллий-маданий хусусиятлари: филол. фанлари докт. дисс. – Тошкент, 2016. – 248 б.; Ганиева Ш.А. Ўзбек фразеологизмлари структураси (шаклий ва мазмуний моделлаштириш): филол. фанлари бўйича фалсафа доктори (PhD) дисс. – Фарғона, 2017. – 155 б.

teglashning lingvistik asos va modellari D.Axmedova tadqiqotlarida o‘rganilgan<sup>16</sup>. Ammo o‘zbek tili frazeologik birliklarini tglashning lingvistik va dasturiy ta‘minoti masalasi biror bir monografik tadqiqotning obyektini bo‘lmagan.

**Dissertatsiya tadqiqotining dissertatsiya bajarilgan oliy ta‘lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog‘liqligi.** Dissertatsiya Toshkent davlat o‘zbek tili va adabiyoti universiteti ilmiy-tadqiqot rejasining “O‘zbek tili birliklariga avtomatik ishlov berish, o‘zbek tili korpusini ishlab chiqish va morfologik tglash istiqbollari” mavzusiga muvofiq bajarilgan.

**Tadqiqotning maqsadi** korpusda o‘zbek tili frazeologik birliklarni tglashning lingvistik va dasturiy asoslarini ishlab chiqishdan iborat.

**Tadqiqotning vazifalari:**

jahon kompyuter lingvistikasida frazeologizmlarni tglash tajribasini o‘rganish, o‘zbek tilshunosligida frazeologizmlarning lisoniy xususiyatlari va nutqiy qiymatini o‘rganish natijalarini umumlashtirish hamda til korpusida iboralarni tglashning ahamiyatini asoslash;

o‘zbek tili frazeologik birliklarini tokenlash va lemmalash, ma‘lumotlar bazasini shakllantirish algoritmlarini ishlab chiqish, shuningdek, o‘zbek tili frazeologik birliklarini morfologik tglashning lingvistik yondashuvlarini aniqlash;

frazeologizmlarni lemmalash algoritmidagi statistik va qoidalarga asoslangan usullarni qo‘llash;

iboralarni morfologik tglashda ma‘lumotlar bazasi, statistik usulni qo‘llash algoritmini aniqlash hamda frazeologizmlarida omonimiyani aniqlash va bartaraf etish texnikasini taklif etish.

**Tadqiqotning obyektini** o‘zbek tili frazeologik birliklari lug‘atlari<sup>17</sup>da berilgan frazeologik birliklar tashkil etadi.

**Tadqiqotning predmetini** o‘zbek tili frazeologik birliklarini tglashning lingvistik asoslari va dasturiy ta‘minotini ishlab chiqish tashkil qiladi.

---

<sup>16</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2018. – 250 б.; Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2019. – 140 б.; Бегматова Г. Ўзбек миллий корпусида идиомалар базасини яратиш: филол. фан. бўйича фалсафа доктори (PhD) дис. автореф. – Термиз, 2021. – 48 б.; Абдурахмонова Н. Электрон корпусларнинг компьютер моделлари: филол. фан. док. дисс. автореф. – Тошкент, 2021. – 74 б.; Тоирова Г.И. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари: филол. фан. док. дисс. автореф. – Бухоро, 2021. – 76 б.; Абдуллаева О. Ўзбек тилининг интернет ахборот матнлари корпусини шакллантиришнинг назарий ва амалий асослари: филол. фан. бўйича фал. док. (PhD) дисс. автореф. – Тошкент, 2022. – 54 б.; Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: филол. фан. бўйича фал. док. (PhD) дисс. автореф. – Термиз, 2021. – 53 б.; Рахманова А. Ўзбек тили Миллий корпусини яратишда компьютер усуллари: филол. фан. бўйича фал. док. (PhD) дисс. – Фарғона, 2022. – 52 б.; Элова Д. Ўзбек тили корпуси бирликларининг услубий тегларини яратиш тамойиллари ва лингвистик таъминот: филол. фан. бўйича фал. док. (PhD) дисс. автореф. – Тошкент, 2022. – 65 б.; N.G‘ulomova. Alisher Navoiy mualliflik korpusi va uning semantik tglari bazasini yaratish (“Badoye” ul-vasat” devoni asosida): filol. fan. bo‘yicha fal. dok. (PhD) diss. avtoref. – Toshkent, 2023. – 52 b.; Xonnazarov E. O‘zbek tili korpusi uchun zamonni ifodalovchi grammatik shakllarni annotatsiyalash: filol. fan. bo‘yicha fal. dok. (PhD) diss. – Toshkent, 2023. – 54 b.; Xusainova Z.Y. O‘zbek tili birliklarini tokenlash, stemlash, lemmalashning lingvistik asoslari va dasturiy ta‘minoti. Filol. fan. bo‘yicha falsafa doktori (PhD) diss. Toshkent, 2024. – 128 b.; Axmedova D.B. Til korpuslarida o‘zbek tili birliklari semantik kengaytmasining leksikografik bazalari: filol. fan. dok. diss. avtoref. – Buxoro, 2024. – 89 b.

<sup>17</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – 407 б.; О‘zbek tili frazeologik lug‘ati [Matn]: lug‘at / Sh.Rahmatullayev, N.Mahmudov, Z.Xolmanova, I.O‘razova, K.Rixsiyeva. – T.: G‘afur G‘ulom nomidagi nashriyotmatbaa ijodiy uyi, 2022. – 636 b.

**Tadqiqotning usullari.** Tadqiqotda statistik tahlil, komponent tahlil, tavsiflash va tasniflash usullaridan foydalanildi.

**Tadqiqotning ilmiy yangiligi** quyidagilardan iborat:

til korpuslari asosida olib borilgan tadqiqotlar natijasida frazemalarning teglash tajribalari tahlil qilinib, o‘zbek tili frazeologik birliklarining o‘ziga xos lisoniy xususiyatlari aniqlanib, ularning tarkibiy qismlari, ularning so‘z turkumlari bilan bog‘liqligi, qat’iy yoki o‘zgaruvchan shakllari, kontekstdagi holatlari va avtomatlashtirilgan tanib olish imkoniyatlari aniqlangan;

til korpusida frazemalarni teglashning dolzarbligi amaliy jihatdan asoslab berilgan, o‘zbek tili frazeologik birliklarini morfologik jihatdan teglashga oid lingvistik yondashuvlar aniqlanib, o‘zbek tilidagi frazemalarni lemmalashga yo‘naltirilgan maxsus algoritm ishlab chiqilgan va frazeologik birliklar ma’lumotlar bazasi shakllantirilgani sababli frazemalarni avtomatik qayta ishlash va ularni lingvistik resurslarga integratsiya qilish imkoniyatlari dalillangan;

frazemalarni lemmalashga doir statistik va qoidalarga asoslangan usullar ishlab chiqilib, frazeologik birliklarni aniqlash va ularni strukturaviy tahlil qilish uchun maxsus morfologik teglar tizimi yaratilgan va n-gram modellari asosida frazemalarni aniqlashga yo‘naltirilgan statistik yondashuv samaradorligi amaliy jihatdan asoslangan;

o‘zbek tili korpusida n-gram usulida frazeologik birlikning grammatik variantlari aniqlanib, frazemalarda omonimiyani aniqlash va bartaraf etish algoritmi, qoidalari ishlab chiqilgan hamda o‘zbek tili korpusiga qo‘llangan.

**Tadqiqotning amaliy natijalari** quyidagilardan iborat:

o‘zbek tili frazeologik birliklari teglar tizimi ishlab chiqilgan;

o‘zbek tili korpusida mavjud frazemalar teglangan;

o‘zbek tili frazemalari ma’lumotlar bazasi shakllantirilgan.

**Tadqiqot natijalarining ishonchliligi** qo‘llanilgan yondashuv, model va algoritm, usul va nazariy ma’lumotlarning ishonchli ilmiy manbalardan olinganligi, keltirilgan tahlillar ratsional, kompleks tahlil metodlari vositasida asoslanganligi, nazariy fikr va xulosalarning amaliyotga joriy etilganligi, olingan natijalarning vakolatli tuzilmalar tomonidan tasdiqlanganligi bilan izohlanadi.

**Tadqiqot natijalarining ilmiy va amaliy ahamiyati.** Dissertatsiyaning ilmiy ahamiyati bir necha tokendan iborat, bir leksik ma’no anglatuvchi birlik – frazemalarni aniqlash algoritmi o‘zbek tili korpusi frazeologik birliklarini avtomatik tanib olish va teglash, til korpusida lisoniy birliklar, jumladan, frazemalarni aniqlash, frazemalarni lemmalash, POS teglash va ularni matnda yaxlit birlik sifatida tanib olish yondashuvlarida ko‘rinishi bilan belgilanadi.

Tadqiqotning amaliy ahamiyati o‘zbek tili korpusi frazeologik birliklarini avtomatik tanib olish, lemmalash, teglash jarayonini ketma-ketligini taqdim etgani, o‘zbek tili korpusi birliklarini avtomatik teglashning umumiy tamoyillarini ishlab chiqishda ma’lumotlar bazasi sifatida xizmat qilishi hamda korpus lingvistikasi yo‘nalishdagi yangi tadqiqotlar uchun nazariy material vazifasini o‘tashi, ilmiy natijalar kompyuter lingvistikasi yo‘nalishi fanlari o‘quv dastur va qo‘llanmalarini

ishlab chiqishda, leksikologiya faniga oid o‘quv qo‘llanmalarni ishlab chiqishda nazariy asos vazifasini bajarishi bilan izohlanadi.

**Tadqiqot natijalarining joriy qilinishi.** O‘zbek tili frazemalari tizimini teglash jarayonida erishilgan ilmiy natijalar asosida:

til korpuslari asosida olib borilgan tadqiqotlar natijasida frazemalarning teglash tajribalari tahlil qilinib, o‘zbek tili frazeologik birliklarining o‘ziga xos lisoniy xususiyatlarini aniqlash, ularning tarkibiy qismlari, so‘z turkumlari bilan bog‘liqligi, qat’iy yoki o‘zgaruvchan shakllari, kontekstdagi holatlari va avtomatlashtirilgan tanib olish imkoniyatlariga oid ilmiy xulosalardan Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti huzuridagi Davlat tilida ish yuritish asoslarini o‘qitish va malaka oshirish markazi davlat muassasasi faoliyatiga joriy qilingan (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti Davlat tilda ish yuritish asoslarini o‘qitish va malakasini oshirish markazining 2025-yil 4-fevraldagi dalolatnomasi). Natijada ta’lim muassasalarida elektron ta’limni joriy etish markazi tasarrufida mavjud elektron resurslardan foydalangan holda o‘zbek tilidagi teglangan 1800 frazeologik birlikdan iborat ma’lumotlar bazasi shakllantirilgan;

til korpusida frazemalarni teglashning dolzarbligi amaliy jihatdan asoslab berilgan, o‘zbek tili frazeologik birliklarini morfologik jihatdan teglashga oid lingvistik yondashuvlar, o‘zbek tilidagi frazemalarni lemmalashga yo‘naltirilgan maxsus algoritm va frazeologik birliklar ma’lumotlar bazasi frazemalarni avtomatik qayta ishlash hamda ularni lingvistik resurslarga integratsiya qilish imkoniyatlariga oid ilmiy xulosalardan O‘zbekiston Milliy universitetida 2022-2024-yillarda bajarilgan REP-25112021/113-raqamli “UzUDT: O‘zbek tilida tabiiy tilni qayta ishlash uchun universal bog‘liqlik daraxti korpusi va uning semantik tahlili” mavzusidagi ilmiy loyihada foydalanilgan (Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti universiteti tomonidan taqdim etilgan 2025-yil 23-apreldagi № 04/11-5273-son ma’lumotnomasi). Natijada o‘zbek tili frazeologik birliklarini lemmalashni amalga oshirish uchun ishlab chiqilgan matematik modellar, algoritmlar va dasturiy majmua asosida gap tarkibidagi frazeologik birliklarni to‘g‘ri aniqlash imkoniyati yaratilgan; semantik konkordans usullari va algoritmlari asosida teglangan 1800 ta frazeologik birlikning sintaktik tahlili aniqlik darajasini 92 foizga yetkazishga erishilgan;

frazemalarni lemmalashga doir statistik va qoidalarga asoslangan usullar, frazeologik birliklarni aniqlash va ularni strukturaviy tahlil qilish uchun maxsus morfologik teglar tizimi va n-gram modellari asosida frazemalarni aniqlashga yo‘naltirilgan statistik yondashuvlarga oid ilmiy xulosalardan Toshkent davlat o‘zbek tili va adabiyoti universitetida 2021-2023-yillarda bajarilgan IZ-2021020813-raqamli “O‘zbek tili ta’limiy korpusini yaratish” mavzusidagi amaliy loyihasida foydalanilgan (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti tomonidan taqdim etilgan 2024-yil 24-dekabrda № 04/1-4250-son ma’lumotnomasi). Natijada o‘zbek tili frazemalarini lemmalash algoritmi va frazemalarning ma’lumotlar bazasi o‘zbek tili ta’limiy korpusi qidiruv tizimini

shakllantirishda qo'llanilgan, shuningdek, frazemalardagi omonimiyani aniqlash va bartaraf etishda algoritmdan samarali foydalanishga erishilgan.

o'zbek tili korpusida n-gram usulida frazeologik birliklarning grammatik variantlarini aniqlash, frazemalardagi omonimiyani aniqlash va bartaraf etish algoritmi hamda unga oid qoidalar Toshkent davlat o'zbek tili va adabiyoti universitetida 2021–2023-yillarda bajarilgan IZ-2021020813-raqamli “O'zbek tili ta'limiy korpusini yaratish” mavzusidagi amaliy loyihasida foydalanilgan (Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti tomonidan taqdim etilgan 2024-yil 24-dekabrda № 04/1-4250-son ma'lumotnomasi). Natijada o'zbek tili ta'limiy korpusi qidiruv tizimida n-gram qidiruv imkoniyati mukammallashtirilgan.

**Tadqiqot natijalarining aprobatsiyasi.** Dissertatsiya natijalari 11 ta ilmiy-amaliy anjumanda, jumladan, 8 ta xalqaro va 3 ta respublika ilmiy-amaliy anjumanida muhokamadan o'tkazilgan.

**Tadqiqot natijalarining e'lon qilinganligi.** Dissertatsiya mavzusi bo'yicha jami 26 ta ish, jumladan, O'zbekiston Respublikasi Oliy attestatsiya komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda 13 ta maqola, jumladan, 10 tasi respublika hamda 3 tasi xorijdagi ilmiy jurnallarda nashr etilgan.

**Dissertatsiyaning tuzilishi va hajmi.** Dissertatsiya kirish, uch asosiy bob, umumiy xulosa, foydalanilgan adabiyotlar ro'yxati va ilovalardan tashkil topgan bo'lib, umumiy hajmi 147 sahifani tashkil etadi.

## DISSERTATSIYANING ASOSIY MAZMUNI

Dissertatsiyaning kirish qismida dissertatsiya mavzusining dolzarbligi va zarurati asoslangan, maqsad va vazifalari, shuningdek, tadqiqotning obyekt va predmetlari tavsiflangan, tadqiqot ishining O'zbekiston Respublikasi fan va texnologiyalarni rivojlantirishning ustuvor yo'nalishlariga mosligi ko'rsatilgan, ilmiy yangiliklar va tadqiqotning amaliy natijalari bayon etilgan, olingan natijalarning ilmiy-amaliy ahamiyati ochib berilgan, tadqiqot ishini amaliyotga joriy qilish, chop etilgan ishlar hamda dissertatsiya tuzilishi va hajmi haqida ma'lumotlar keltirilgan.

Tadqiqotning **I bobi “O'zbek tili frazeologik birliklarning umumiy tavsifi”** deb nomlanadi. Bobning birinchi paragrafida *jahon frazeologiyasi va korpus lingvistikasi munosabati* tahlilga tortiladi.

Jahon frazeologiyasida turli tillarning frazeologik birliklari har xil aspektda tadqiq etilgan. Frazeologik tadqiqotlar korpus lingvistikasi uchun turli darajada nazariy manba vazifasini bajaradi. Birinchidan, bu tadqiqot natijalari til korpuslarida frazeologik birliklarni teglash (annotatsiyalash) uchun nazariy asos vazifasini o'tasa, boshqa tomondan, bunday tadqiqotlar kompyuter leksikografiyasida ahamiyatli: bu tadqiqot natijalari asosida turli tillarning frazeologik elektron lug'atlarini yaratishga zamin hozirlaydi.

Lingvistik korpuslar va frazeologiyaning munosabati ikki jihatda yaqqol ko'rinadi:

birinchidan, korpusga asoslangan frazeologik tadqiqotlar;  
ikkinchidan, frazeologik birliklarini til korpuslarida aks ettirish;  
uchinchidan, frazeologiyani o'rganishda korpusdan lingvodidaktik vosita sifatida foydalanish.

Ilmiy adabiyotlarda korpus bilan ishlashning quyidagi asosiy turlari ajratiladi:

- korpusga yo'naltirilgan tadqiqotlar (corpus-driven);
- korpusga asoslangan tadqiqotlar (corpus-based);
- korpusni illyustrativ manba sifatida ishlatadigan tadqiqotlar va material (corpus-illustrated, corpus-informed)<sup>18</sup>.

Ushbu turlar orasidagi chegara ma'lum tadqiqotlarda buzilsa-da, korpusga asoslangan yondashuvlarda geterogen tadqiqotning mazkur usuli yaxshi natija beradi, chunki korpusga yondashilgan tahlilda har bir holat alohida tahlil qilinadi. Demak, tadqiqotchi korpusdan illyustrativ material manbasi sifatida foydalansa, (corpus-illustrated yondashuv), uning vazifasi, birinchi navbatda, paradigmatic misollar kontekstini tanlashdir. Bu tamoyil, xususan, leksikografik jarayonlarda dolzarblik kasb etadi. Korpusdan illyustrativ material olingan lug'at lisoniy bilim/hisga asoslangan yoki leksikografning fayllar kabinetida tasodifan topilgan illyustrativ misollar asosida tuzilgan lug'atga qaraganda nisbatan obyektivligi bilan ajralib turadi. D.O.Dobrovolskiyning fikriga ko'ra, illyustrativ material bilan ishlashning korpusga asoslangan tamoyili leksikografni har bir holatda alohida qaror qabul qilishdan ozod qiladi, degani emas. Aksincha, bu tamoyilning afzalligi shundaki, u tanlash imkoniyatlarini kengaytiradi va obyektiv ma'lumotlarga asoslanib, muayyan shakldan foydalanish darajasini aniqlashga imkon beradi<sup>19</sup>.

Jahon korpus lingvistikasidagi frazeologiyaga oid qator tadqiqotlar korpus vositasida yoki korpus asosida olingan natijalar tahlili asosida dunyoga kelgan. Bu kabi tadqiqotlarda til korpuslari va unda aks ettirilgan frazeologik birliklar lingvistik material yoki tadqiqot metodi (korpus tahlil metodi) vazifasini o'taydi. Bu kabi tadqiqotlarda frazeologik birliklar tahlili birinchi planda bo'ladi, til korpuslari shunchaki tadqiqot vositasiga aylanadi.

Lingvistik korpuslar va frazeologiya bilan bog'liq tadqiqotlarning yana bir ko'rinishi borki, ular til korpuslarida frazeologik birliklarini teglash, frazemalarni korpusda aks ettirish yoki ularning qidiruv tizimini ishlab chiqish, frazemalarni izlash va tahlil qilishda til korpuslaridan foydalanishga bag'ishlanadi. Bunday tadqiqotlarda korpus tuzish, uni ishlatish, lingvistik korpusda frazeologik birliklarni aks ettirish muammosi birinchi planga chiqadi.

Jahon tilshunosligida frazeologik birliklarning korpusga asoslangan tadqiqot usullari ko'p qo'llaniladi. Bu usul tahlilning asosli ekanligi, tahlil materiali – korpus materiallarining obyektivligi bilan ahamiyatli. Tahlilning bunday usuli tadqiqot natijalarining aniqligini kafolatlaydi. Ko'plab maqolalarni tahlil qilish turli leksik va frazeologik birliklarning korpusga asoslangan tahlili ommalashganligini ko'rsatadi. Ammo til korpuslarida frazeologik birliklarni teglash muammosi kam o'rganilgan

<sup>18</sup> Добровольский Д.О. (2020). Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов. Вестник Санкт-Петербургского университета. Язык и литература, 17(3). – С. 398–411. <https://doi.org/10.21638/spbu09.2020.303>

<sup>19</sup> Добровольский Д.О. Ко'rsatilgan manba. – С. 399-400.

soha ekanligi ko‘rinadi. Lingvistik korpuslarda frazeologik birliklarning teglanishi korpus asosida tadqiq olib borish imkoniyatlarini yanada kengaytiradi.

Bobning ikkinchi paragrafida *o‘zbek tilshunosligida frazeologizmlarning o‘rganilishi, lisoniy xususiyatlari va nutqiy qiymati* tahlil qilinadi.

Frazeologik birliklar til leksikasining ajralmas qismi bo‘lib, ma‘lum ma‘noni ifodalashga yordam beradi, tasvirlaydi va nutqning ekspressivligini oshiradi. Iboralar, shuningdek, lingvomadaniy jihatdan o‘ziga xos bo‘lib, boshqa tillarda ularning ekvivalentlari bo‘lmasligi mumkin. Ma‘lum bir tilning frazeologik birliklari jamiyatning lisoniy muloqoti va madaniy taraqqiyoti jarayonida shakllanadi, rivojlanadi. Frazeologik birliklarning shakllanishi va rivojlanishiga ta‘sir qiluvchi omillar sifatida quyidagilarni ajratish mumkin: (1) tarixiy kontekst, (2) madaniy xususiyatlar, (3) lisoniy jarayon, (4) metafora va obrazlar, (5) boshqa tillarning ta‘siri. Tilning frazeologik birliklarining shakllanishi va rivojlanishi jamiyatning lisoniy va madaniy evolyutsiyasini aks ettiruvchi dinamik jarayondir. Ular tilning muhim qismi bo‘lib, nutqning boyligi va ifodaliligiga hissa qo‘shadi.

O‘zbek tili frazeologizmlarining o‘rganilishida tuzilgan turli lug‘at va o‘quv qo‘llanmalar ahamiyatli. O‘zbek tili frazeologik birliklarining xususiyatlari maqolalarda turli darajada o‘rganilgan.

Frazeologizmlarning ekspressivlik xususiyati, frazeologizmlar tarjimasini masalalari, matbuot tilida okkazional frazeologizmlar, o‘zbek tilining izohli lug‘atida frazeologiya muammolari, shaxs tavsifli frazemalarning semantik tahlili, so‘zlashuv nutqi frazemalarining mavzuiy guruhlari, frazeolekt va lug‘atshunoslik munosabati, badiiy nutqda frazeologizmlar, o‘zbek va afg‘on tillarida o‘xshash fe‘l frazeologik birliklar, frazeologizmlarning adabiy til normasiga munosabatiga doir tadqiqotlar o‘zbek tili korpuslarida frazeologik birliklarning turli xususiyatlarini belgilash, teglash va ma‘lumotlar bazasiga kiritishda muhim material bo‘lib xizmat qiladi.

Birinchi bobning uchinchi paragrafida *til korpusida iboralarni teglashning lingvistik asoslari va ma‘lumotlar bazasi* masalasi tahlil qilingan. Frazeologik birliklarni korpusda teglash, til korpuslarining lingvistik ma‘lumotlar bazasida frazeologik birliklarni yetarli axborot bilan ta‘minlash uchun ularning leksik-semantik, morfologik, grammatik, uslubiy va boshqa xususiyatlarini belgilab olish maqsadga muvofiq. O‘zbek tili frazeologik birliklari bu jihatdan atroflicha tadqiq etilgan. Korpusda frazeologik birliklarni teglash uchun mavjud ilmiy material – ma‘lumotlarni tizimlashtirish bizning ishimiz uchun yetarli bo‘ladi. Frazeologik **birlik**larga grammatik nuqtayi nazardan qaraganda, iboralarning ayrimi bir so‘zga teng kelib, bir ma‘noni anglatadi. Boshqasi so‘z birikmasiga teng bo‘ladi; ba‘zilari gapga teng kelib, nisbiy tugal fikrni bildiradi. Mana shu xususiyatlariga ko‘ra, frazeologik **birlik**lar grammatik jihatdan dastlab uchga bo‘linadi (Qarang: ilovadagi 1-chizma). Shundan kelib chiqib, teglar sirasiga frazeologik birliklarning grammatik belgisiga ko‘ra so‘z-ibora, birikma-ibora, gap-ibora teglari kiritiladi.

Frazeologizmlar gapda har qanday bo‘lak bo‘lib kela oladi:

a) ega: *Otaxonning yuziga oyoq qo‘yish bizning odatimizda yo‘q (hurmatsizlik ma‘nosida)* (A.Muxtor.);

b) kesim vazifasida: *Shunaqa vaqtda odamning yuragi qinidan chiqib ketadi* (qo‘rqadi ma’nosida) (A.Qahhor.);

d) aniqlovchi: *O‘rinbosarning bunday xushomadlari oldida og‘zining tanobi qochadigan* odati bor edi (iljayadigan ma’nosida) (Gazetadan.);

e) to‘ldiruvchi: *Qo‘rqar edi sho‘rlik ko‘zni yumishdan* (o‘lishdan ma’nosida) (H.Olimjon.);

f) hol: *Hozir dunyoni boshiga ko‘tarib buvim kelar* (dodlab, mojaro ko‘tarib ma’nosida).

Frazeologik birliklar sintaktik qurilishiga ko‘ra so‘z birikmasi, sodda va qo‘shma gap shakllarida keladi (Qarang: ilovadagi 1-chizma). Frazeologik birliklar tuzilishiga ko‘ra turini aniqlash ularni va lemmalash teglash jarayoni uchun muhim.

Tadqiqotning II bobi “O‘zbek tili korpusida frazemalarning statistik tahlili” deb nomlangan. Birinchi paragraf o‘zbek tili frazemalarining morfologik shakli va variantlari tadqiqiga an’anaviy va korpus tahlili yondashuviga bag‘ishlangan.

Variantlanish – barcha til birliklariga, shu jumladan, frazeologik birliklarga ham mansub hodisa. Ko‘p variantli iboralarning asosiy qismini ikki va uch variantli iboralar tashkil etadi. Umuman, frazeologik birliklarda variantlanish boshqa til birliklariga nisbatan kuchli bo‘lib, bir iboraning variantlari soni o‘ntagacha yetadi. Variantlanish tufayli bir ibora boshqa bir iboraga aylanmaydi, bir iboraga xos obraz birligi, lug‘aviy ma’no birligi saqlanadi. Variantlanish iboraga xos yaxlitlikning o‘zi doirasida, shu yaxlitlik yo‘l qo‘ygan darajada voqe bo‘ladi. Aks holda, yo ibora yo‘qqa chiqadi, yoki boshqa bir ibora barpo qilinadi.

Frazeologik variantlarda leksik va grammatik variantlar farqlanadi. Iboraning tarkibida mustaqil turkumdagi so‘z bilan ifodalangan komponentni o‘zgartirib hosil qilingan variantga *leksik variant* deyiladi. Bunda so‘z komponent: almashtiriladi, qo‘shiladi, tashlanadi. Bulardan eng murakkabi leksik almashtirish asosidagi variantlanishdir<sup>20</sup>.

Ibora tarkibini grammatik jihatdan o‘zgartirish bilan hosil bo‘ladigan variantga *grammatik variant* deyiladi. Iboraning tashqi grammatik qurilishiga (iborani kontekstga singdirish bilan) bog‘liq grammatik o‘zgarishlar variantlanish deb qaralmaydi. Jumladan, kontekstga singdirish munosabati bilan ot iboraning sonda, kelishikda o‘zgarishi, fe‘l iboraning tuslovchi, egalik olib o‘zgarishi, mayl, zamon yasovchilarini olib o‘zgarishi variantlanish emas, balki *paradigmatik formasi* deyiladi<sup>21</sup>. Iboraning bunday varianti, aniqrog‘i, paradigmatik shakli tilda cheksiz bo‘ladi. Istalgan iborani til korpuslaridan qidirsak, uning yuzlab grammatik varianti yoki morfologik shakli uchraydi. Masalan: “achchig‘i kelmoq” iborasi O‘zbek tili korpusidan qidirilganda uning quyidagi paradigmatasi aniqlandi:

achchig‘i keldi	achchig‘i kelgan
achchig‘i kelib	achchig‘i kelyapti
achchig‘im keladi	achchig‘i keladi-da
achchig‘ini keltiradigan	achchig‘i keladi
achchig‘i kelur	achchig‘im kelgan

<sup>20</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

<sup>21</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

achchiqlangan kelinni  
achchig'ini keltirdi  
achchig'ing keladi  
achchig'i kelsa

achchig'imni keltirdi  
achchig'lari kelgan-da  
achchig'i kelgandek

Frazeologik birlik yuqoridagi tip paradigmatic shakllarida biror chegaralanishga ega bo'lsa, shunda grammatik variantlanish o'rinli. Masalan, fe'l iboraning bo'lishli yoki bo'lishsiz shaklda kelishi (*ko'kka sovurdi va ko'kka sovurmadi*) umumiy holat bo'lib, grammatik variantlanish deb qaralmaydi. Shu bilan bir vaqtda, faqat bir aspektda ishlatiladigan, mavjud bo'ladigan iboralar ham bor. Masalan, *alifni tayoq deya olmaslik* iborasi odatda bo'lishsiz aspektda ishlatiladi, ba'zan bo'lishli aspektda uchrab qoladi: *alifni tayoq deydigan* kabi. Endi buni frazeologik variant deb tan olish kerak. Tom ma'nodagi grammatik variantlar iboraning ichki grammatik qurilishida (iboraning kontekst bilan aloqador bo'lmagan qismlarida) sodir bo'ladigan o'zgarishlar natijasida yuzaga keladi. Jumladan, fe'l iboralarda nisbat (daraja) yasalishi va shu bilan bog'liq holda yuz beradigan grammatik o'zgarishlar shunday variantlanish turlaridan biridir. Masalan, *adab(i)ni bermoq* fe'l iborasi tarkibidagi fe'l so'z-komponent aniq darajada, tushum kelishigidagi so'z-komponent bilan to'ldiruvchili birikma hosil etib turadi. Xuddi shu ibora *adab(i) berildi* shaklida ham ishlatiladi. Nisbatning o'zgarishi iboraning ichki sintaktik qurilishini keskin o'zgartirib yuboradi: birikmaga teng holat gapga teng holatga o'tadi va shunga bog'liq boshqa qator grammatik-semantik o'zgarishlar yuz beradi<sup>22</sup>.

Grammatik variantlanish ham, xuddi leksik variantlanish kabi, asosan ibora tarkibidagi biror grammatik qismni almashtirish, tushirib qoldirish yoki unga qo'shish bilan sodir bo'ladi. Grammatik variantlanishda o'ziga xos hodisa – so'z-komponentlarning joylashish tartibi asosida yuzaga keladigan variantlanish ham mavjud:

**1) almashtirish asosida paydo bo'ladigan variantlanish:** *O'tirishga joy topolmaslik // o'tirgani joy topolmaslik* (bu yerda maqsad ravishdoshi yasovchisi -**gani** harakat nomi yasovchi -**sh** va jo'nalish kelishigi shakli bilan almashgan); *qo'yn(i)ni puch yong'oqqa to'ldirmoq // qo'ynini puch yong'oq bilan to'ldirmoq* (bu yerda kelishik shakli ko'maksi bilan almashgan);

**2) qo'shish asosida paydo bo'ladigan variantlanish:** *holdan toymoq // holdan toydirmoq* (orttirma daraja qo'shilgan); *ko'ngl(i) bo'shadi // ko'ngl(i) bo'shashdi* (harakatning biroz kuchsiz voqe bo'lishini ifodalovchi affiks qo'shilgan);

**3) tushirib qoldirish asosida variantlashish:** *ko'z(i) ilindi, ko'z ilindi* (egalik shakli tushib qolgan)<sup>23</sup>.

Yuqoridagi ma'lumotlardan ko'rinib turibdiki, har bir iboraning variantini topish va uni lug'atga kiritish uchun uning nutqiy qo'llanishini katta dataset (aniqrog'i, til korpusi) asosida kuzatish kerak. Buni qo'lda amalga oshirish imkoni mavjud emas. Katta ma'lumotlar bilan ishlashga mo'ljallangan korpuslarda

<sup>22</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

<sup>23</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

фразеологик birlikmalarni teglash orqali har bir iboraning grammatik varianti yoki morfologik shaklini topishga imkon yaratiladi.

Фразеологик variantlilik hodisasiga berilgan quyidagi ta'riflarni solishtiramiz:

1. "Фразеологик variantlar bu invariant фразеологик birlikning ma'no tarkibi va mazmunidagi, stilistik va sintaktik vazifalari farq qiluvchi turli ko'rinishidir, invariantning qisman leksik yoki morfologik o'zgarishi, so'zshakli yoki so'z tartibi bilan farq qiluvchi ko'rinishidir"<sup>24</sup>.

2. Idiomaning struktur varianti deb tarkibi va tuzilishida jiddiy farq bo'lmagan фразеологик birliklarga aytiladi<sup>25</sup>.

3. Фразеологик variantlar uslubiy va bo'yoqdorlik bilan farqlanib, turli kontekstda ularning maqbul varianti qo'llanadi<sup>26</sup>.

Yuqoridagi ta'riflarni umumlashtirgan holda aytish mumkinki, фразеологик variantlar bir invariantning nutqiy ko'rinishlari bo'lib, ular nutq uslubi hamda buyoqdorligiga qarab farqli matnlarda turli variantlarda uchraydi.

Y.P.Ignatovich, T.N.Fedulenkovalar esa tilda фразеологик variantlilikning mavjudligini ta'kidlagan holda, iboralarda variantlilikning quyidagi turilarini farqlashadi:

1. Leksik variativlik.
2. Leksik-morfologik variativlik.
3. Kvantativ variativlik.
4. Leksik-kvantativ variativlik.
5. Sintaktik variativlik.
6. Leksik-sintaktik variativlik.
7. Sintaktik-kvantativ variativlik<sup>27</sup>.

Shuningdek, mualliflar maqolada ingliz tili frazemalarida eng ko'p variantlilik foizi leksik variantlilik ekanligini aniqlashgan<sup>28</sup>.

Tilshunoslikda shu vaqtga qadar фразеологик birliklarning variantlarini aniqlashda korpus yondashuviga asoslangan tadqiqotni uchratmadik. Aslida, til korpuslarida frazemalar teglangan bo'lsa yoxud n-gram funksiyasi mavjud bo'lsa, frazemalarning grammatik, leksik varianti hamda grammatik paradigmasini aniqlash mumkin; bu tahlil natijalari yanada aniqroq ekanligi bilan ishonchli sanaladi. Quyida o'zbek tilida korpusi qidiruvi natijasi hamda o'zbek tilining фразеологик lug'ati<sup>29</sup>dagi ayrim iboralarning morfologik shakllari va grammatik variantlarini tahlil qilamiz.

<sup>24</sup> Кунин А.В. Основные понятия английской фразеологии как лингвистической дисциплины и создание англо-русского фразеологического словаря. Дисс. д-ра филол. наук. – М. 1964. – С. 442.

<sup>25</sup> Амосова Н.Н. Основы английской фразеологии. – Л., 1963. – С. 100.

<sup>26</sup> Авалиани Ю.Ю., Ройзензон Л.И. О разграничение синонимии и вариантности в области фразеологических единиц // Вопросы фразеологии и составления фразеологических словарей. – Баку, 1968. – С. 72.

<sup>27</sup> Игнатович Я.П., Федуленикова Т.Н. Вариант фразеологической единицы как показатель ее тождества (на материале фразеологии с компонентом – глаголом отчуждения) // [https://journals.tsu.ru/uploads/import/1583/files/37\\_035.pdf](https://journals.tsu.ru/uploads/import/1583/files/37_035.pdf)

<sup>28</sup> Игнатович Я.П., Федуленикова Т.Н. Вариант фразеологической единицы как показатель ее тождества (на материале фразеологии с компонентом – глаголом отчуждения) // [https://journals.tsu.ru/uploads/import/1583/files/37\\_035.pdf](https://journals.tsu.ru/uploads/import/1583/files/37_035.pdf)

<sup>29</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – 407 б.

Frazeologik lug‘atda *abjag‘ini chiqarmoq* iborasining 1 ta morfologik shakli ko‘rsatilgan: *abjag‘i chiqdi*<sup>30</sup>. Ammo o‘zbek tili korpusi qidiruvi natijasida mazkur iboraning quyidagi 10 ta morfologik shakli aniqlandi:

<i>abjag‘i chiqqan</i>	<i>abjag‘imni chiqarib</i>
<i>abjag‘i chiqib</i>	<i>abjag‘ini chiqar</i>
<i>abjag‘ing chiqqan</i>	<i>abjag‘ini chiqarganga</i>
<i>abjag‘i chiqdi</i>	<i>abjag‘ini chiqargan</i>
<i>abjag‘ingni chiqaraman</i>	<i>abjag‘ini chiqaradi</i>

Olingan natijalardan ko‘rinadiki, bu iboraning ikkala komponentida ham morfologik shakllar o‘zgarib qo‘llanishi mumkin. “abjag‘(i)” qismi turli egalik va kelishik shakllarini, “chiqarmoq” shakli esa zamon, sifat-dosh, ravishdosh shakllarida qo‘llangan. Ibora turli morfologik shakllarda uchragan bo‘lsa-da, grammatik variant hosil qilmagan.

“Andaza olmoq” iborasi esa frazeologik lug‘atda keltirilmagan. Lug‘atning kirish qismida “mazkur lug‘atda faqat yuqori chastotali yoki sinonimik, antonimik juftlari, variantlari mavjud ibora keltirilgan deyiladi”<sup>31</sup>. Ammo korpusdagi qidiruv natijasi ko‘rinadiki, o‘zbek tilida “andaza olmoq” iborasi quyi chastotali emas. Korpusda bu iboraning quyidagi morfologik shakllari aniqlandi:

<i>andaza olish</i>	<i>andaza olgin</i>
<i>andaza olgan</i>	<i>andaza oladi</i>
<i>andaza olsa</i>	<i>andaza oladigan</i>
<i>andaza olgandek</i>	<i>andaza olinayotganidan</i>
<i>andaza olingan</i>	

Shu asosda mazkur iborani ham frazeologik lug‘atga kiritish maqsadga muvofiq.

#### *O‘zbek tili korpusi yordamida frazemalar chastotasini aniqlash*

Korpusda idioma yoki frazemalarni aniqlash oddiy qidiruv bilan topibmaydi. Chunki n-gram usulida frazemalar qidiruvini amalga oshirganda natijalar orasiga ma’lum bir tilda ibora sanalmaydigan so‘z qo‘shilmalari ham ro‘yxatga kirib qoladi. n-gram usulini frazemalar qidiruvi uchun eng optimal usul deb hisoblasak-da, natijalarni qo‘lda tahlil qilish, natijalar sirasini tozalash (ibora bo‘lmagan birliklarni ro‘yxatdan o‘chirish) jarayoni, albatta, bajariladi. Masalan, “baxti ochilmoq” iborasi qidiruvi quyidagi natijani berdi:

- |                         |                      |
|-------------------------|----------------------|
| 1) baxtim ochilmadi     | 6) baxting ochilsin  |
| 2) baxti ochilishi      | 7) baxtiqaro oqsochi |
| 3) baxting ochilmadi-da | 8) baxtidan ochish   |
| 4) baxti ochilmayapti   | 9) baxti ochilsa-ku  |
| 5) baxti ochilib        |                      |

Keltirilgan ro‘yxatdan ko‘rinib turibdiki, mazkur ibora korpusda 9 marta uchragan yoki “baxt” va “ochilmoq” so‘zlari shuncha marta uchragan. Ammo bu ro‘yxatdagi “baxtiqaro oqsochi” iborasi mazkur paradigmaga mansub emas. Demak,

<sup>30</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 22.

<sup>31</sup> Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

o‘zbek tili korpusida n-gram usuli qidiruvda xato so‘z qo‘shilmasini olib tashlaganda, “baxti ochilmoq” iborasi 8 chastotaga ega.

O‘zbek tili frazeologik qatlami juda boy bo‘lib, ular turli darajada qo‘llanadi: ayrim frazemalar faol qo‘llansa, ba’zi frazemalarning qo‘llanish chastotasi past. O‘zbek tili frazemalarining chastotasini o‘zbek tili korpusi<sup>32</sup> vositasida aniqlash uchun dastlab frazemalar POS teglangan bo‘lishi talab etiladi. Bunday qidiruvni amalga oshirish uchun dastlab o‘zbek tili frazemalari ma’lumotlar bazasi shakllantirildi, frazemalar lug‘at shakliga keltirildi, lug‘at shakli iboraning lemma shakli deb tan olindi, shundan keyin ot-ibora, sifat-ibora, fe‘l-ibora sifatida teglandi (bu haqda keyingi bobda batafsil yoziladi).

Korpusdagi frazemalar lemmalangan bo‘lsa, korpusda frazemalar qidiruvini amalga oshirish mumkin. Shunda past va yuqori chastotali birliklar aniqlanadi. Quyidagi 1-jadvalda o‘zbek tilidagi ba’zi frazemalarning korpus qidiruvdagi chastotasini taqdim etamiz (Qarang: 1-jadval).

<i>No</i>	<i>Frazema</i>	<i>Frazema anglatgan ma’nosiga ko‘ra turkumi</i>	<i>chastotasi</i>
1	abjag‘i chiqqan	<i>Fe‘l-ibora</i>	10
2	aksiga olib	<i>Fe‘l-ibora</i>	2
3	andaza olish	<i>Fe‘l-ibora</i>	9
4	aql bovar qilmaydigan	<i>sifat-ibora</i>	26
5	bahri-dilim ochildi	<i>Fe‘l-ibora</i>	17
6	barmog‘ini tishlamoq	<i>Fe‘l-ibora</i>	

### **1-jadval. O‘zbek tili frazemalari qo‘llanish chastotasiga uzschoolcorpora.uz korpusidan misollar**

1-jadvaldan ma’lum bo‘ladiki, o‘zbek tilida barcha frazemalarning ham qo‘llanishi bir xil emas. Tanlab olingan 5 ta frazema orasida belgi ma’nosini ifodalovchi “*aql bovar qilmaydigan*” (26 marta qo‘llanish), “*bahri-dili ochilmoq*” frazemalari chastotasi nisbatan yuqori, “*aksiga olib*” frazemasining chastotasi past ekanligi aniqlandi. Bu xulosa, albatta, nisbiy. Chunki biz natijalarni o‘zbek tilining ta’limiy korpusidan oldik, u korpus, asosan, maktab darsliklari, maktab yoshiga mos badiiy asarlar hamda internet sayti xabarlarini matni o‘rin olgan. Bunday matnlarda frazemalarning kamroq uchrashi oddiy hol.

Tadqiqotning **III bobi “O‘zbek tili frazeologik birliklarini teglashning dasturiy asoslari”** deb nomlangan. Birinchi paragrafda o‘zbek tili korpusi asosida *frazeologik birliklar ma’lumotlar bazasini shakllantirish* masalasi yoritilgan.

Lingvistik ma’lumotlar bazasi so‘z, ibora, grammatik shakl, idioma va boshqa lisoniy unsurlarni saqlaydigan ma’lumotlarning tartiblangan to‘plamidir. Bunday ma’lumotlar bazalaridan turli maqsadlarda foydalanish mumkin, jumladan:

1. Leksikografiyada: lug‘at va ma’lumotnomalar yaratish.

<sup>32</sup> <https://uzschoolcorpora.uz/>

2. Tabiiy tilni qayta ishlash: avtomatik tarjima, nutqni aniqlash va matn generatsiyasi bilan bog‘liq texnologiyalarni qo‘llab-quvvatlash.

3. Lingvistik tadqiqotlar: til tuzilishi, til birliklarining o‘zgarishi va qo‘llanilishini tahlil qilish.

4. Ta’limda: til o‘rganish uchun o‘quv materiallari va vositalarini yaratish.

Ma’lumotlar bazasi ma’lum bir maydonlarga ega jadvallardan tuzilishi mumkin, bu yerda har bir til birligining ta’rifi, gap bo‘laklari, illyustrativ misollar va boshqa metama’lumotlarni qamrab oladi.

Lingvistik ma’lumotlar (til birliklari)ning ma’lumotlar bazasini yaratish bir necha bosqichda amalga oshiriladi. Bu kabi ma’lumotlar bazasini yaratishning bosqichma-bosqich qadamlarini quyidagicha izohlaymiz:

**1. Maqsad va talab belgilanadi.** Ma’lumotlar bazasini tuzishdan oldin ma’lumotlar bazasi nima uchun ishlatilishi aniqlanadi:

- MB qaysi til birliklarini (so‘z, ibora, grammatik qurilmalar) o‘z ichiga olishi belgilab olinadi;

- har bir birlik haqida qanday ma’lumotlar saqlanadi (ta’rif, so‘z turkumi, misollar, sinonim, antonim va boshqalar);

- ma’lumotlarni qidirish va filtrlash mezonlari aniqlanadi.

**2. Platformani tanlash.** Ma’lumotlar bazasini saqlash uchun platforma tanlanadi va belgilanadi:

- relyatsion ma’lumotlar bazalari (masalan, MySQL, PostgreSQL, SQLite).

- NoSQL ma’lumotlar bazalari (masalan, MongoDB).

- kichik loyihalar uchun elektron jadval muharrirlari (masalan, Microsoft Excel, Google Sheets).

**3. Ma’lumotlar bazasi tuzilishini loyihalash.** Ma’lumotlar bazasi uchun jadval tuzilishi yaratiladi. Asosiy jadvallar quyidagilardan iborat bo‘lishi mumkin:

- til birliklari jadvali: har bir birlik haqidagi asosiy ma’lumotlarni o‘z ichiga oladi;

- izoh(ma’no)lar jadvali: ta’rif, izoh va qo‘llanish holatlari misollari tavsiflanadi;

- grammatik belgilar jadvali: so‘z turkumi, shaxs, son va boshqalar grammatik belgilar haqida ma’lumot.

**4. Ma’lumotlarni kiritish.** Ma’lumotlarni yig‘ish va jadvallarga kiritish jarayoni qo‘lda bajaridishi mumkin yoki skriptlar yordamida avtomatlashtirilishi mumkin (masalan, matn ma’lumotlarini tahlil qilish uchun Python yordamida).

**5. Kirish interfeysini ishlab chiqish.** Agar ma’lumotlar bazasi boshqa foydalanuvchilar tomonidan foydalanilsa, qulay foydalanish uchun interfeys yaratish ma’lumotga muvofiq. Bunday interfeysni quyidagi ko‘rinishda yaratish mumkin:

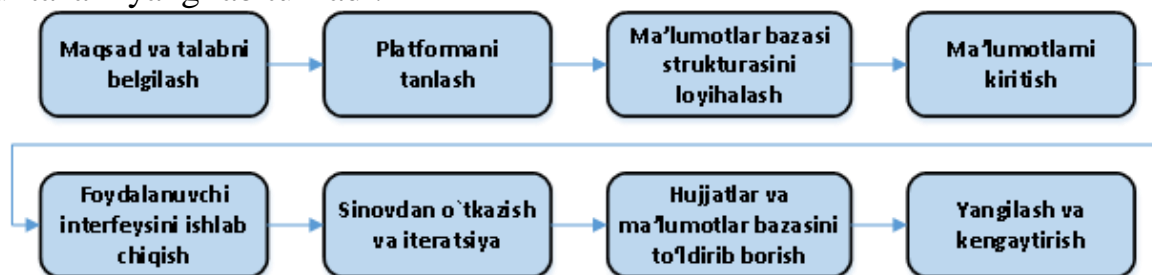
- ma’lumotlarni kiritish va izlash shakllariga ega veb-ilova;

- mobil ilova.

**6. Sinov va iteratsiya.** Ma’lumotlar bazasining maqsadga muvofiq ishlashiga ishonch hosil qilish uchun sinovdan o‘tkazish talab etiladi. Foydalanuvchi(lar)ning fikr-mulohazalari asosida xatolar tuzatiladi.

**7. Hujjatlar va ma'lumotlar bazasini to'ldirib borish.** Ma'lumotlar bazasidan foydalanishda hujjatlar, yangi til birliklari va tuzatishlar qo'shish orqali yangilab turiladi.

**8. Yangilash va kengaytirish.** Yangi birliklar qo'shish, xatolarni tuzatish va foydalanuvchi ehtiyojlariga ko'ra strukturani yaxshilash orqali ma'lumotlar bazasi muntazam yangilab turiladi.



**2-chizma. Ma'lumotlar bazasini yaratishning bosqichlari**

Yuqorida ko'rsatilgan bosqichlarni bajarish orqali funksional va samarali lingvistik ma'lumotlar bazasini yaratishga erishish mumkin.

Yuqorida lingvistik ma'lumotlar yaratishning umumiy tavsifi keltirildi. Quyida frazeologik ma'lumotlar bazasi haqida fikr yuritimiz.

Frazeologik ma'lumotlar bazasi leksikografik manbalarni boyitish, til o'qitish va lingvistik tadqiqotlar uchun xizmat qiladi. Lingvistik ma'lumotlar bazasining bu ko'rinishini tuzishda ayrim o'ziga xos tomonlarni e'tiborga olish lozim.

### **1. Ma'lumotlar bazasi tarkibiga qo'yilgan talablar:**

1. Frazeologik birliklarning tipi (ibora, maqola, turg'un birikma)ni aniqlash.
2. Atributlarni aniqlash:
  - frazeologizm (ibora);
  - izohi;
  - tarjimasi (agar ma'lumotlar bazasi boshqa til ma'lumotlarini ham qamrab olsa);
  - misollar;
  - mavzu doirasi, kategoriyasi: masalan, tabiat, oziq-ovqat va h.;
  - qo'llanilish darajasi;
  - sinonim va antonimlari (eagar mavjud bo'lsa).

### **3. Ma'lumotlar bazasi tarkibini loyihalash talab etiladi:**

Ma'lumotlar modeli: ma'lumotlar bazasi tuzilishini shakllantirish. Masalan:

- "iboralar" jadvali (ID, frazeologizm, izohi, kategoriyasi)
- "misollar" jadvali (ID, frazeologizm ID, misol)
- "tarjimalar" jadvali (ID, frazeologizm ID, tarjima, tili)

Lingvistik korpusdan foydalangan holda frazeologik birliklar ma'lumotlar bazasini yaratish tildan olingan real misollar asosida frazeologik birliklarni to'plash, tahlil qilish va ma'lumotlar bazasini tuzishning samarali usuli hisoblanadi. Bu jarayon quyidagi algoritm asosida amalga oshadi:

**1. Lingvistik korpusni tanlash.** Bu jarayonda ma'lumotlar bazasi uchun manba bo'ladigan korpus tanlanadi. Turli xil matnli materiallarni (masalan, yangilik, badiiy adabiyot, ilmiy maqola) o'z ichiga oladigan lingvistik korpus tanlash

maqsadga muvofiq. Frazеologik birlik (idiomani oson qidirish va chiqarib olish uchun maxsus korpus dasturiy ta'minotidan (masalan, Sketch Engine, AntConc) foydalanish mumkin.

**2. Frazеologik birliklarni ajratib olish.** Frazеologik birliklarni qidirish (turg'un iboralarni topish) uchun tanlangan korpusda qidirish va filtrlash funksiyalaridan foydalanish talab etiladi (masalan, "fe'l + ot" qolipidagi birikmalarni qidirish orqali).

### 3. Qidiruvni avtomatlashtirish.

- skriptlar va algoritmlar: agar obyekt katta hajmdagi matn bo'lsa, oddiy iboralar yoki so'zni qayta ishlashning boshqa usullaridan foydalangan holda iboralarni avtomatik ravishda ajratib olish uchun Python yoki boshqa dasturlash tilidagi skriptlar orqali qidiruvni amalga oshirish mumkin;

- mashinali o'qitish: katta hajmdagi matnlardan idiomalarni tasniflash va ajratib olish uchun mashinali o'rganish usullaridan foydalanish yaxshi natija beradi.

Bular frazeologik birliklar ma'lumotlar bazasini tuzishning umumiy algoritmi va tavsifi bo'lib, mazkur paragrafda o'zbek tili korpusi asosida frazeologik birliklarning ma'lumotlar bazasini ishlab chiqishning aniq algoritmiga to'xtalamiz.

Oldingi paragraflarda aytilganidek, tadqiqot obyekti o'zbek tili korpusi bo'lib, frazeologik ma'lumotlar bazasi mazkur korpus asosida shakllantirildi. Buning uchun quyidagi ketma-ketlikda so'rov berish natijasida har bir frazeologik birlikning ro'yxati grammatik paradigmasi topildi, keyinchalik shular asosida ma'lumotlar bazasi shakllantirildi.

O'zbek tili korpusi ma'lumotlar bazasidan frazeologik birliklarni aniqlashda quyidagi ER sxemadan foydalanamiz (Qarang: 2-chizma).

Bu paragrafda turli chastotali frazeologik birliklar ma'lumotlar bazasidan qanday o'rin egallashi, ma'lumotlar bazasining tuzilishi va tarkibi, o'zbek tili frazeologik birliklarining statistikasi masalasini tahlil qilamiz.

```
SELECT dbo.Token.Id, dbo.Token.Value, dbo.Token.CorrectVersion,  
dbo.WordToken.LemmaValue, dbo.WordToken.LemmaSuffixAffsPart,  
dbo.WordToken.StemValue, dbo.WordToken.StemPrefixAffsPart,  
dbo.WordToken.StemSuffixAffsPart, dbo.WordToken.RootValue
```

```
FROM dbo.Token INNER JOIN
```

```
dbo.WordToken ON dbo.Token.Id = dbo.WordToken.TokenId
```

```
WHERE dbo.WordToken like % "ajali yet(moq)"
```

Masalan, "ajali yetmoq" iborasini topish uchun quyidagi algoritmda so'rov amalga oshiriladi:

1. Quyidagi havolani ochish kerak:  
<https://uznatcorpara.uz/uz/Bigram/Search>

The image shows a web search interface with the title "Bigram bo'yicha izlash". It contains several input fields and dropdown menus. The first row has two text input fields labeled "1-so'z" and "2-so'z", followed by a dropdown menu labeled "Like" and a button labeled "Izlash". The second row has two dropdown menus and another dropdown menu labeled "Harf registri inobatga olinmasin".

## 1-rasm. Frazologik birliklar qidiruv oynasi

Ushbu interfeysda 2 ta token (“ajali” va “yet(moq)”)ni mos kataklarga kiritish lozim.

2. “Harf registri inobatga olinsin/olinmasin” qismidan mos elementni tanlash kerak.

3. Korpus ma’lumotlar bazasidan yuqorida keltirilgan soʻz birikmasini mos yozuvlarni “Like” (%) yoki “Regulyar ifoda”lardan foydalanib izlash mumkin.

4. Zarur izlash parametrlari oʻrnatilgach, “Izlash” tugmasini bosish lozim.

5. Natija quyidagi shaklda hosil qilinadi (Qarang: ilovadagi 3-chizma):

Natijada, “ajali yetmoq” iborasining quyida berilgan grammatik paradigmasi hosil boʻldi:

- |                    |                       |
|--------------------|-----------------------|
| 1. ajaling yetgani | 8. ajalimiz yetganida |
| 2. ajali yetgan    | 9. ajali yetgani      |
| 3. ajali yetib     | 10. ajali yetar       |
| 4. ajali yetgudek  | 11. ajalim yetguncha  |
| 5. ajalim yetib    | 12. ajaling yetmay    |
| 6. ajali yetmagan  | 13. ajali yetganini   |
| 7. ajali yetganda  |                       |

Roʻyxatdan koʻrinib turganidek, mazkur morfologik shakllar unikal boʻlib, biri birini takrorlamaydi. Hosil qilingan paradigmada xato morfologik shakl yoki variant mavjud emas. Bundan koʻrinib turibdiki, morfologik ma’lumotlar bazasiga “ajali yetmoq” iborasining 13 ta morfologik shakli kiritiladi.

Yuqorida keltirilgan algoritm asosida *aksiga olmoq*, *anqoning urugʻi*, *alamidan chiqmoq*, *aqli kirmoq* kabi iboralarning paradigmasi aniqlandi (Qarang: ilovadagi 2-jadval). 3-jadvaldan koʻrish mumkinki, 01.04.2024 holatidagi 51890123 gapdan tashkil topgan korpusda barcha frazeologik birliklarning qoʻllanish chastotasi bir xil emas: *aksiga olmoq* iborasi 2 marta, *anqoning urugʻi* iborasi 5 marta (shundan bittasi iboraning sinonimi boʻlib chiqdi), *alamidan chiqmoq* 17 marta, *aqli kirmoq* 23 marta ishlatilgan deb roʻyxat avtomatik shakllandi.

Ma’lumotlar bazasini shakllantirish jarayonini ushbu ketma-ketlikda quyidagi biznes-jarayonda ifodalaymiz (Qarang: ilovadagi 4-chizma).

Demak, ma’lumotlar bazasida mazkur frazeologik birlikning birinchi (original) shakli, morfologik paradigmasi, varianti va sinonimini kiritish mumkin.

01.04.2024 holatidagi 51890123 gapdan tashkil topgan korpusda *aqli kirmoq* iborasining 23 ta kollokatsiyasi aniqlangan. Mazkur kollokatsiyalar orasida *aqliy oʻtkirlik*, *aqlining oʻtkirligi* kabi kollokatlar frazeologik birlik boʻlolmaydi, yaʼni bu iboraning varianti emas. Shu sababli tozalash jarayonida bu kollokatlar oʻchiriladi, natijada, ma’lumotlar bazasiga *aqli kirmoq* iborasining 21 ta morfologik varianti kiritiladi. Mazkur iboraning varianti lugʻatda koʻrsatilgan boʻlgani, u leksik variant (bitta soʻzi butunlay boshqa) boʻlgani uchun uning variantining morfologik paradigmasini shakllantiramiz.

Korpus qidiruvida *es(i) kirmoq* iborasining morfologik paradigmasi quyidagicha: *esi kirganidan*, *esini kiritib*, *esingiz kirarli-chiqarli*, *esi kirdi-chiqdi*, *esini kiritish*, *esi kirib*, *esimiz kirmaydi*, *esi kirsas*, *esi kirmagan*, *esim kiradigan*, *esi*

*kirgan, esing kirmadi, esingni kirgizib, esi kirsin, esing kirdi, esi kirib, esimni kiritib, esi kirar-chiqar, esi kirari-chiqari, esing kiradi, esini kirgizib, esi kirmasdanoq, esini kiritolmasalar-da, esi kirarli-chiqarli, esi kiradi, esimizni kiritmoqchi.*

### ***n*-gram usuli haqida**

*n*-gram usuli turg'un lisoniy birliklarni izlash va tahlil qilish uchun optimal vosita bo'lib xizmat qilishi mumkin. Biroq aniq natijaga erishish va korpusda bunday birliklar qidiruv natijasini ko'rish aniqligini oshirish uchun boshqa usul va yondashuvlarni birgalikda qo'llash muhim. Frazeologik birliklar bazasini yaratish uchun lingvistik korpusdan foydalanish nafaqat tildagi frazemalarning yorqin namunalarini aniqlaydigan, balki til o'qitish va tadqiqot uchun foydali resurs yaratadigan kuchli va zamonaviy yondashuvdir.

*n*-gram usuli turg'un lisoniy birliklar, masalan, frazeologik birliklar, idioma va aforizmlar, tilda tez-tez qo'llaniladigan boshqa iboralarni topishda juda samarali vositadir.

Korpusdagi frazeologik birliklarni aniqlashda *n*-gramdan foydalanish chastota tahlili asosida turg'un iboralarni aniqlash imkonini beruvchi kuchli va samarali usuldir. Ammo shuni esda tutish kerakki, *n*-gram modellari ko'pincha frazeologik birliklarni ajratib olishning aniqligini oshirish uchun qo'shimcha ishlov berish va tahlil qilishni talab qiladi.

## **XULOSA**

1. Til korpusi bilan ishlashning (1) korpusga yo'naltirilgan; (2) korpusga asoslangan; (3) korpusni illyustrativ manba sifatida ishlatadigan tadqiqotlar kabi turlari ajratiladi. Korpus lingvistikasida til korpuslarida aks etgan frazeologik birliklar lingvistik material yoki tadqiq metodi vazifasini o'taydi; frazeologik birliklar tahlil obyekti bo'lib, til korpuslari shunchaki tadqiq vositasi vazifasini bajaradi. Til korpuslarida frazeologik birliklarini teglash, frazemalarni korpusda aks ettirish, frazemalarning qidiruv tizimini ishlab chiqish, frazemalarni izlash va tahlil qilishda til korpuslaridan foydalanishga oid tadqiqotlar ham ommalashgan.

2. Frazema – ikki va undan ortiq leksemaning semantik-sintaktik birlashuvidan tarkib topgan, shaklan birikma yoki gapga teng bo'lgan, ko'chma ma'noli turg'un (barqaror) qurilma. Frazemalarni korpusda teglash, til korpusining lingvistik ma'lumotlar bazasida frazemalarni yetarli axborot bilan ta'minlash uchun ularning leksik-semantik, morfologik, grammatik, uslubiy va boshqa xususiyatlarini belgilab olish talab etiladi.

3. Grammatik variantlanish ibora tarkibidagi biror grammatik qismni almashtirish, tushirib qoldirish yoki unga qo'shish bilan sodir bo'ladi. Tilda iboraning varianti/paradigmatik shakli cheksiz bo'ladi. Til korpuslarida bir iboraning yuzlab grammatik varianti yoki morfologik shakli uchraydi. Har bir iboraning variantini topish va uni lug'atga kiritish uchun uning nutqiy qo'llanishini katta dataset (aniqrog'i, til korpusi) asosida kuzatish kerak. Katta ma'lumotlar bilan ishlashga mo'ljallangan korpuslarda frazeologik birikmalarni teglash orqali har bir iboraning grammatik varianti yoki morfologik shaklini topishga imkon yaratiladi.

4. Ayrim iboralarning ikkala komponentida ham morfologik shakllar o'zgarib qo'llanishi mumkin. Ba'zan ibora turli morfologik shakllarda uchrasa-da, grammatik

variant hosil qilmaydi. Bunday iboralarning varianti mavjud emas. Ayrim iboralar frazeologik lugʻatda keltirilmagan, ammo oʻzbek tili korpusida bu iboraning katta hajmli morfologik shakllari aniqlandi. Shu sababli oʻzbek tili izohli lugʻatida berilmagan iboralarni til korpusidan aniqlash oʻzbek tili frazeologik lugʻatni boyitishga xizmat qiladi.

5. Oʻzbek tili frazeologik qatlami juda boy boʻlib, ular turli darajada qoʻllanadi: ayrim frazemalar faol qoʻllansa, baʼzi frazemalarning qoʻllanish chastotasi past. Oʻzbek tili frazemalarining chastotasini oʻzbek tili korpusi vositasida aniqlash uchun dastlab frazemalar POS teglash talab etiladi. Bunday qidiruvni amalga oshirish uchun dastlab oʻzbek tili frazemalari maʼlumotlar bazasi shakllantirildi, frazemalar lugʻat shakliga keltirildi, lugʻat shakli iboraning lemma shakli deb tan olindi; ot-ibora, sifat-ibora, feʼl-ibora sifatida teglandi.

6. Lingvistik maʼlumotlar bazasi soʻz, ibora, grammatik shakl, idioma va boshqa lisoniy unsurlarni saqlaydigan maʼlumotlarning tartiblangan toʻplamidir. Maʼlumotlar bazasi maʼlum bir maydonlarga ega jadvallardan tuziladi; har bir til birligining taʼrifi, gap boʻlaklari, illyustrativ misollar va boshqa metamaʼlumotlarni qamrab oladi. Lingvistik maʼlumotlar bazasini yaratish maqsad va talabni belgilash; platformani tanlash; maʼlumotlar bazasi strukturasi loyihalash; maʼlumotlarni kiritish; kirish interfeysini ishlab chiqish; sinov; jujjatlar va maʼlumotlar bazasini toʻldirib borish; yangilash va kengaytirish kabi bosqichlarda amalga oshiriladi.

7. Maʼlumotlar bazasini shakllantirish jarayonini iboraning morfologik paradigmasini shakllantirish; mazkur iboraning variantlarini aniqlash; iboraning variantini lugʻatdan tekshirish; agar ibora lugʻatda mavjud boʻlmasa, lugʻatga kiritish; mazkur iboraning sinonimini aniqlash (agar mavjud boʻlsa); maʼlumotlar bazasida bir ibora satrini yakunlash ketma-ketligidagi jarayonda amalga oshadi. Maʼlumotlar bazasida frazeologik birliklar korpusdagi chastotasiga qarab, past chastotali, oʻrta chastotali hamda yuqori chastotali frazeologik birliklar guruhiga boʻlinadi.

8. Oʻzbek tilida frazeologik birliklar orasidagi omonimiya hamda frazeologik birlik va erkin birikma orasidagi omonimiya mavjud. Qidiruv natijasida roʻyxatda ikkala birlik aralash tarzda turadi. Tahlil va tadqiqot uchun bu roʻyxatni qoʻlda qayta tahlil qilishga ehtiyoj tugʻiladi. Bu masalani yechishda n-gram modellarini qoʻllash mumkin.

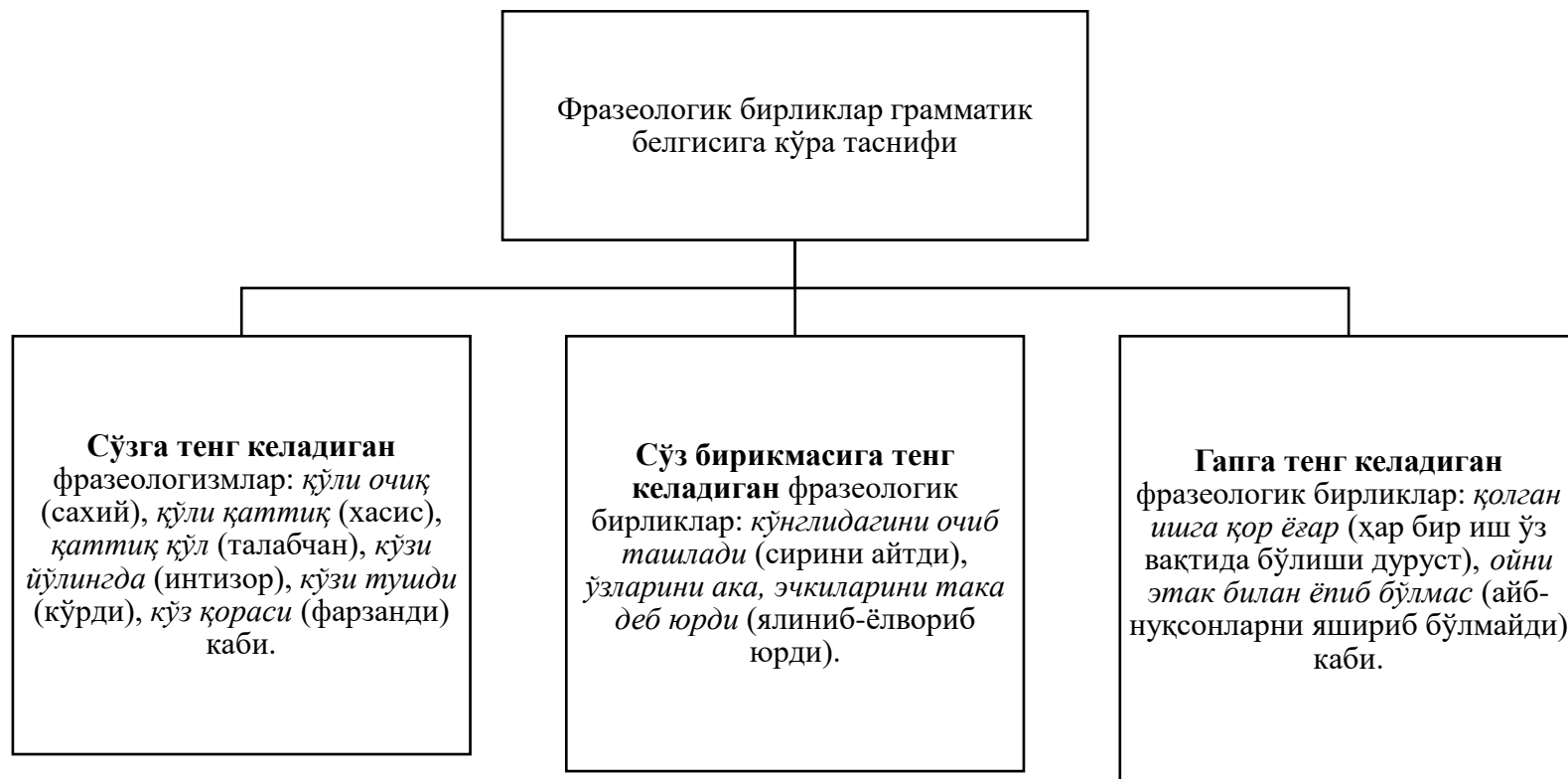
9. n-gram usuli turgʻun lisoniy birliklar: frazeologik birlik, idioma, aforizm, tilda tez-tez qoʻllaniladigan boshqa iboralarni topishda samarali vositadir. n-gram yordamida turgʻun lisoniy birliklarni aniqlash n-gram hosil qilish; chastotani tahlil qilish; n-gramlarning tipik uzunligini aniqlash; turgʻunlik va kontekstni aniqlash; filtrlash; klasterlash; kontekstni qayta ishlash; qoʻlda tekshirish va ekspertiza; turli misollarda qoʻllash jarayonlaridan oʻtadi.

10. Lugʻatdagi frazemalarni korpusdan qidirish usuli samarali natija koʻrsatadi: maʼlum iboralarni aniq topish va ularning qoʻllanish chastotasini hisoblash imkoniyatini beradi. Morfologik normallashtirish va bir necha variantlarni qamrab olish (sinonim shakllar, feʼl zamonlari, shaxs-son qoʻshimchalari) orqali frazemalarni yuqori aniqlikda topish mumkin. Eng mos yondashuv sifatida lugʻatga asoslangan qidiruv usuli, zarur hollarda qoʻshimcha statistik filtrlar usuli tavsiya etiladi. Lugʻat bazasini qoʻllash bilan birga, korpusdan mustaqil ravishda

frazemalarni aniqlash uchun n-gram frekvensiya tahlili, PMI, t-test kabi statistik metodlar va ML usullarini qo'llash ham istiqbolli yo'nalishdir. Bu usullar yangi, lug'atlarda qayd etilmagan barqaror birikmalarni aniqlashga yordam beradi.

11. Tadqiqot natijasida korpusda eng ko'p va eng kam ishlatiladigan frazeologizmlar ro'yxati, ularning uslubiy xususiyatlari aniqlandi. Bu ma'lumotlar frazeologik belgilar bilan korpusni teglash ishlariga zamin yaratdi. Jumladan, avtomatik frazema taggeri yordamida korpusdagi har bir frazeologik birikmani maxsus [IB] tegi bilan teglash mumkin. Bu esa keyinchalik ularni qidirish, filtrlash, tarjima qilishda muhim ahamiyat kasb etadi.

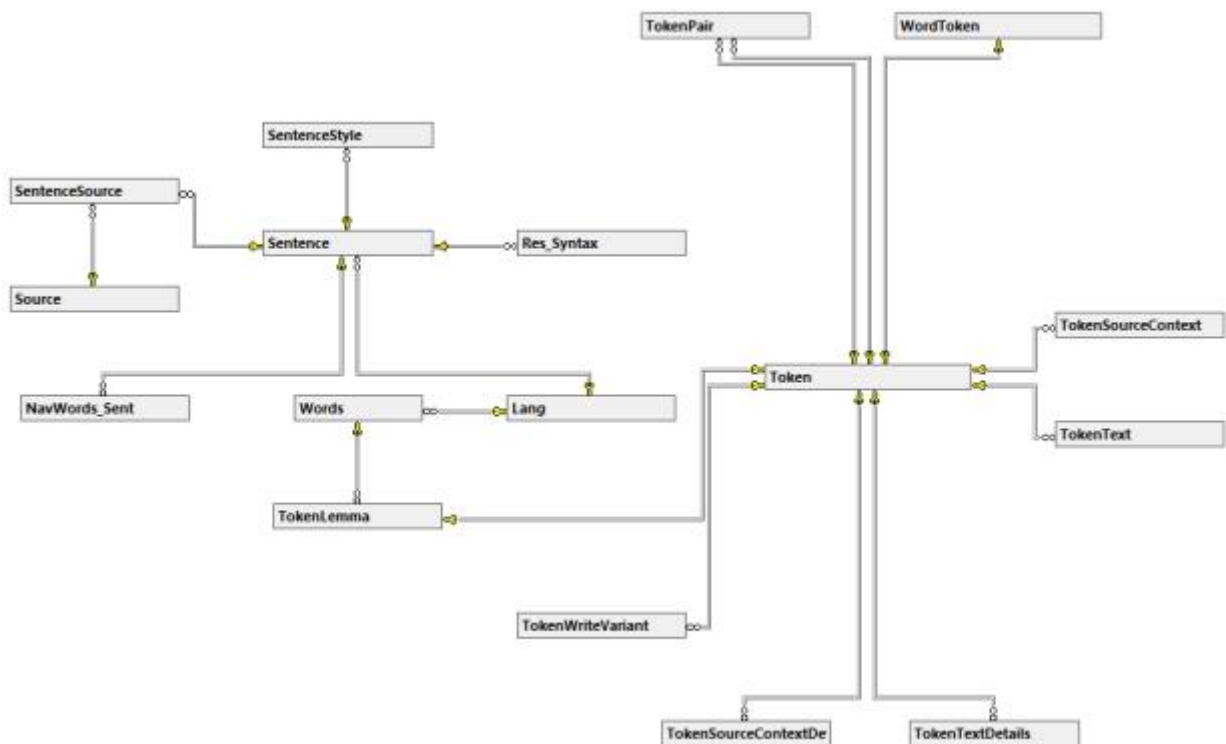
12. Frazemalarni semantik va sintaktik tasniflash, har bir toifaning turli janrlardagi ulushini o'rganish imkoni paydo bo'ladi. Kelgusida frazeologizmlarning soha bo'yicha qo'llanish ko'rsatkichlari ham tadqiq qilinishi mumkin: masalan, siyosiy nutqda keng ishlatiladigan iboralar bilan badiiy adabiyotga xos iboralarni qiyosiy tahlil qilish imkoni mavjud.



**1-chizma. Frazеologik birliklarning grammatik belgisiga ko‘ra tasnifi**

Asabni buzmoq		Avjga chiqmoq	
asabini buzgani-yu	asabim buziladi	avjga chiqqani	avjga chiqayapti
asabini buzib	asabni buzadi	avjga chiqadi	avjga chiqishi
asabim buzilib	asabim	avjga chiqqan	avjga chiqsa
asabini buzishdan	buzilganidanmi-	avjga chiqmoqda	avjga chiqdi
asablarini	yo‘qmi	avjga chiqdi	avjga chiqqanini
buzishlari	asablari buzilgani	avjga chiqishiga	<b>mavjudotga chiqarib</b>
asabni buzadigan	asabingni	avjga chiqqandi	<b>mavjudligi ochiqlandi</b>
asabi buziladi	buzmaslik	avjga chiqqanida	avjga chiqarishdi
asabi buzilgani	asablarim buzildi	avjga chiqib	avjga chiqqanligini
asabini buzadigan	asabim buzildi	avjga chiqaramiz	avjga chiqaveradi
asabimni buzdi	asabi buzilib	avjga chiqadigan	avjga chiqa
asabingizning	asab buzilishidan	avjga chiqarayotgan	avjga chiqaveradi
buzilgani	asabimizni	avjga chiqqanini	zavjalariga chiqarardi
<b>asabiy</b>	buzaverib	avjga chiqqan	Zavjangizga achchiq
<b>buzilishlar</b>	asabi buzildi	avjga chiqyapti	avji chiqib
asab buzilishi	asabi buzilishidan	avjga chiqqanda	avjga chiqqani
asabini buzadi	asabingiz buzilib	<b>mavjudotlar kirib-chiqib</b>	avjga chiqdiki
asab	asabini buzishni	avjga chiqishi	avjga chiqarkan
buzilishining	asabi buzilganligini	avjga chiqqan	avjga chiqardi
asab buzishiga	asab buzilishlari	avjga chiqayotganligidan	<b>Mavjuda chiqib</b>
asab buzilishiga	asabini buzari	avjga chiqarilgan	avjga chiqqanmish
asabingizni	asab buzilishlar	avjga chiqqanda	<b>mavjud chiqindilarni</b>
buzadi	asabini buzdi	avjga chiqishida	avjga chiqqach
asabini buzayapti	asabimni	avjga chiqqaniga	avjga chiqmoqda
asabni	buzganlaring	avjga chiqayotgan	avjga chiqmasa
buzmasdan	asablarimning	avjga chiqargan	avjga chiqqanining
asabi buzilgandan	buzilishi	avjga chiqyapti	<b>mavjlanib chiqqan</b>
asab buzilish	asabing buziladi	avjga chiqadi	<b>mavjudligi chiqindilar</b>
asabingiz	asabingizni	avjga chiqqanidan	avjga chiqarib
buzilishiga	buzishning	avjga chiqar	<b>mavjudligini ochiq</b>
asablari buzilib	asablarni buzishga	avjga chiqa	<b>mavjudot chiqib</b>
asabi buzilgan	asabingni	avjga chiqqanligi	<b>mavjudligi</b>
asabimizni buzib	buzmasdan	avjga chiqib	<b>ochiqchasi</b>
	asabingizni	mavjud chiqindi	<b>mavjud ochiq</b>
	buzmang	avjga chiqar	
	asabini buzayotgan	avjga chiqmagandek	
	asabini buzishga	avjga chiqarayotgan	
	asablar	avjga chiqmoqchi	
	buzilishining	avjga chiqardi	
	asablar buzilishi		
	asablari buzilgan		
	asabini buzish		

2-jadval. Past va yuqori chastotali frazeologik birliklar



2-chizma. frazeologik birliklarni aniqlashda quyidagi ER sxemasi

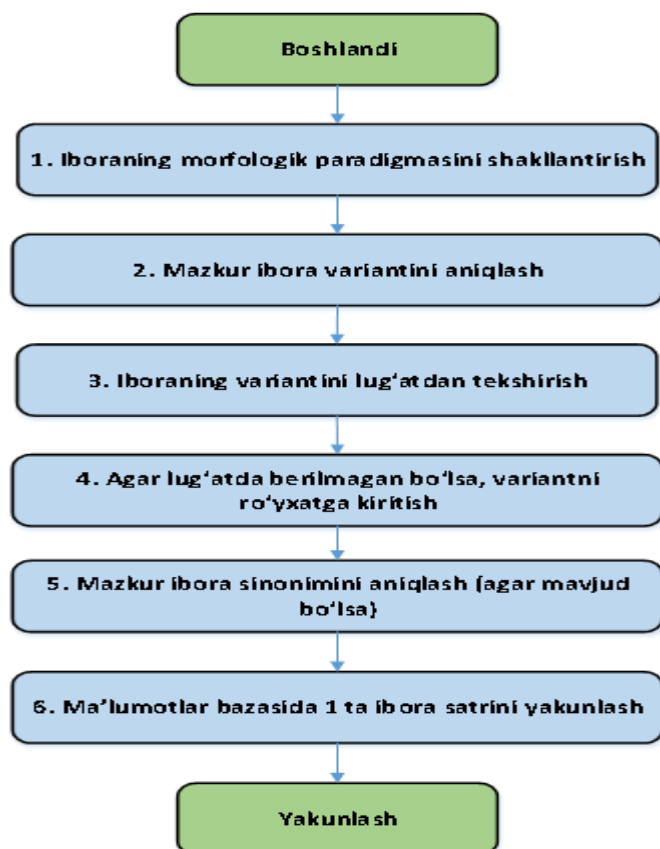
Bigram bo'yicha izlash

Natija: 24 ta yozuvdan 24 tasi ko'rsatilmogda.

**Bigram**

Ajali yetgan
ajaling yetgani
ajali yetgan
ajali yetib
ajali yetgudek
ajalim yetib
ajali yetmagan
ajali yetganda

3-chizma. Frazologik birliklar qidiruv natijasi



**4-chizma. Ma'lumotlar bazasini shakllantirish jarayoni**

**ОДНОРАЗОВЫЙ УЧЕНЫЙ СОВЕТ НА БАЗЕ УЧЕНОГО СОВЕТА  
№ DSc.03/30.12.2019.Fil.19.01 ПО ПРИСУЖДЕНИЮ НАУЧНЫХ  
СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ ГОСУДАРСТВЕННОМ  
УНИВЕРСИТЕТЕ УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ ИМЕНИ  
АЛИШЕРА НАВОИ**

---

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ ИМЕНИ АЛИШЕРА НАВОИ**

**ЁДГОРОВ УМИДЖОН САЙДИЛЛА УГЛИ**

**ЛИНГВИСТИЧЕСКИЕ И ПРОГРАММНЫЕ ОСНОВЫ  
ТЕГИРОВАНИЯ ФРАЗЕОЛОГИЗМОВ В КОРПУСЕ УЗБЕКСКОГО  
ЯЗЫКА**

**10.00.11 – Теория языка. Прикладная и компьютерная лингвистика.**

**АВТОРЕФЕРАТ**  
диссертации доктора философии (PhD) по филологическим наукам

**Toshkent – 2025**

**Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за В2023.3.PhD/Fil3963.**

Диссертация выполнена в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои.

Автореферат диссертации на трёх языках (узбекском, русском, английском (резюме)) размещён на веб-сайте Научного совета ([www.tsuul.uz](http://www.tsuul.uz)) и в Информационно-образовательном портале «Ziyonet» ([www.ziyonet.uz](http://www.ziyonet.uz)).

<b>Научный руководитель:</b>	<b>Элов Ботир Болтаевич</b> доктор философии (PhD) по техническим наукам, доцент
<b>Официальные оппоненты:</b>	<b>Абдурахмонова Нилуфар Зайнабиддин кизи</b> доктор филологических наук (DSc), профессор <b>Эшмуминов Аскар Алламуродович</b> доктор филологических наук (DSc), доцент
<b>Ведущая организация:</b>	<b>Самаркандский государственный университет</b>

Защита диссертации состоится «\_\_\_» \_\_\_\_\_ 2025 года в \_\_\_\_\_ часов на заседании разового Научного совета DSc.03/30.12.2019.Fil.19.01 при Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои (Адрес: 100100, город Ташкент, Яккасарайский район, улица Юсуфа Хас Ходжиба, дом № 103. Тел.: (99871) 281-42-44; факс: (99871) 281-42-44; e-mail: [monitoring@navoiy-uni.uz](mailto:monitoring@navoiy-uni.uz))

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои (зарегистрировано за № \_\_\_\_\_). Адрес: 100100, город Ташкент, Яккасарайский район, улица Юсуфа Хас Ходжиба, дом № 103. Тел.: (99871) 281-42-44; факс: (99871) 281-42-44.

Автореферат диссертации разослан «\_\_\_» \_\_\_\_\_ 2025 года.  
(Реестр протокола рассылки № \_\_\_\_\_ от «\_\_\_» \_\_\_\_\_ 2025 года).

**Х. Дадабоев**  
Председатель \_\_\_\_\_ разового  
диссертационного совета, созданного  
на основе научного совета по  
присуждению учёных степеней, доктор  
филологических наук, профессор

**К.У. Пардаев**  
Учёный секретарь \_\_\_\_\_ разового  
диссертационного совета, созданного  
на основе научного совета по  
присуждению учёных степеней, доктор  
филологических наук, профессор

**С.Б. Нормаматов**  
Председатель научного семинара при  
разовом диссертационном совете,  
созданном на основе научного совета  
по присуждению учёных степеней,  
доктор филологических наук,  
профессор

## **ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))**

**Актуальность и востребованность темы диссертации.** В мировой компьютерной лингвистике при автоматической обработке естественного языка активно изучается проблема лингвистической разметки различных языковых единиц, включая лексемы, математической лингвистике и практике машинного перевода. В разработке программ разметки для различных языков задача разметки фразеологических единиц рассматривается как отдельная проблема. Эти новшества в науке послужили основой для формирования перспективных научных направлений, связанных с применением информационных технологий в лингвистике. Ведь это породило необходимость изучения общих принципов морфологического анализа для информационно-поисковых систем, корпусной лингвистики и машинного перевода, а также разработки систем создания формального языка и грамматики для лингвистического обеспечения анализирующих систем. Это, в свою очередь, расширяет возможности поиска фразеологических единиц в корпусах, распознавания фразеологических единиц как целостных единиц в автоматическом переводе, а также разработки тезаурусов.

В мировой лингвистике в конце XX века на повестку дня были поставлены такие задачи, как лингвистическое моделирование языка, автоматическое распознавание языковых единиц в корпусах, лемматизация фразеологических единиц в рамках отдельных языков, теория и практика тегирования, разработка соответствующих алгоритмов. Автоматическое тегирование фразеологизмов и их выделение в корпусе упрощает анализ. Лемматизация фразем облегчает статистический анализ их употребления в корпусах, что позволяет делать обоснованные выводы о семантике, прагматике и стилистических характеристиках фразем. Это важный шаг на пути создания систем автоматической обработки естественного языка, таких как машинный перевод, сентимент-анализ и извлечение информации. Выделение (тегирование) фразеологических единиц в языковом корпусе способствует изучению фразеологии, включая выявление закономерностей употребления фразем, а также их семантических и структурных особенностей. Корпуса с лемматизированными фраземами используются при создании различных лингвистических ресурсов, таких как фразеологические словари, учебные пособия и программное обеспечение для изучения языков.

В области узбекской компьютерной лингвистики внимание уделяется научным исследованиям в таких направлениях, как языковой корпус, автоматический перевод, искусственный интеллект, понимание и обработка узбекского языка. Однако задача разработки алгоритма и модели автоматического тегирования фразеологических единиц в корпусе узбекского языка не изучена на научно-теоретическом уровне. Ведь «...сохранение чистоты государственного языка, его обогащение и повышение речевой культуры населения; обеспечение активной интеграции государственного языка в современные информационные технологии и коммуникации» является одной из актуальных задач, стоящих сегодня перед узбекской компьютерной

лингвистикой. В этой связи важное значение приобретают такие вопросы, как достижение автоматизированной обработки узбекского языка с помощью современных информационных технологий, создание языковых корпусов, электронных переводчиков, тезаурусов и других первичных автоматических средств обработки, а также обеспечение их лингвистической поддержки.

Указанный текст, подготовленный в соответствии с Указом Президента Республики Узбекистан от 13 мая 2016 года № ПФ-4997 «О создании Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои», Указом от 17 февраля 2017 года № ПФ-2789 «О мерах по дальнейшему совершенствованию деятельности Академии наук, организации, управления и финансирования научно-исследовательских работ», Указом от 21 октября 2019 года № ПФ-5850 «О мерах по коренному повышению престижа и статуса узбекского языка как государственного», Указом от 20 октября 2020 года № ПФ-6084 «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики», Постановлением от 4 октября 2019 года № ПК-4479 «О широком праздновании тридцатилетия принятия Закона Республики Узбекистан “О государственном языке”», Посланием Президента Республики Узбекистан Олий Мажлису от 24 января 2020 года, а также другими нормативно-правовыми актами, служит определённой основой для реализации задач, обозначенных в вышеуказанных документах.

**Соответствие исследований приоритетным направлениям развития науки и технологий Республики.** Исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики - I. “Формирование системы инновационных идей и путей их реализации в развитии информационного общества и демократического государства в социальном, правовом, экономическом, культурном и духовно-просветительском аспектах”.

**Степень изученности проблемы.** В мировой корпусной лингвистике вопросы создания языкового корпуса и автоматической лингвистической разметки корпусных единиц проанализированы в исследованиях Г.И.Кустовой, О.Н.Ляшевской, Ю.В.Падучевой, Ю.В.Рахилиной. Среди работ, посвящённых изучению лингвистических особенностей фразеологических единиц, заслуживают внимания исследования А.Кэтрин<sup>33</sup>, Й.Дегсдоттир<sup>34</sup>, Р.Сариф<sup>35</sup>. Кроме того, на монографическом уровне изучались такие вопросы, как употребление идиоматических единиц в речи<sup>36</sup>, типологические особенности фразеологических единиц английского и чешского языков, а

---

<sup>33</sup> Kathryn A. A Study of Idioms in Relation to Language Universals. A Senior Thesis submitted in partial fulfillment of the requirements for graduation in the Honors Program. – Liberty University: Fall, 2015. – Pages 47.

<sup>34</sup> Ylva Dagsdóttir. The teaching of idiomatic expressions in upper secondary school after the implementation of the LK20, from the perspectives of English language teachers. Master’s Thesis. – Lektorutdanning i språkfag, 2022. – Pages, 92.

<sup>35</sup> Rifka Nur’aini Syarif. An analysis of idiomatic expression meaning in frozen film script and its implication in teaching vocabulary. State institute for islamic studies of metro, 2020. – Pages 127.

<sup>36</sup> Elizabeth Jane Holt. Figures of speech: an exploration of the use of idiomatic phrases in conversation. – The University of York, Department of Sociology. March, 1991. – Pages 291.

также корпусное исследование фразем американского варианта английского языка<sup>37</sup>.

В узбекском языкознании существует ряд исследований, посвящённых языковым особенностям и прагматическому изучению фразеологических единиц. Заслуживают внимания исследования<sup>38</sup> И.Кучкартаева, М.Рузыкуловой, К.Хаитметова, Б.Юлдашева. Ш.Рахматуллаев изучал отдельные вопросы узбекской фразеологии<sup>39</sup>, основные грамматические особенности глагольных оборотов современного узбекского языка, а также сочетаемость глагольных фразем в узбекском языке. Абдугафур Маматов исследовал вопросы формирования узбекских фразеологизмов<sup>40</sup>, Г.Жуманазарова - употребление фразем в речи<sup>41</sup>, Б.Юлдашев - функционально-стилистические особенности фразеологических единиц в современном узбекском литературном языке, М.Юлдашев - мастерство использования народных выражений у Чулпона. А.Маматов и Д.Очилов занимались исследованием синтаксических и деривационных моделей фразем, А.Маматов - вопросами фразеологической стилистики и основ формирования фразеологизмов<sup>42</sup>. Следует также отметить, что Ш.Алматова, М.Вафаева, А.Носиров, Ш.Ганиева исследовали фразеологические единицы в различных аспектах<sup>43</sup>.

---

<sup>37</sup> Rafatbakhsh, E., Ahmadi, A. A thematic corpus-based study of idioms in the Corpus of Contemporary American English. *Asian. J. Second. Foreign. Lang. Educ.* 4, 11 (2019). <https://doi.org/10.1186/s40862-019-0076-4>

<sup>38</sup> Кучкартаев И. Фразеологическое новаторство Абдулла Каххара: Автореф. дис. канд. филол. наук. – Ташкент, 1965. (Кўчқортоев И. А.Қаҳҳорнинг фразеологик маҳорати: филол. фанл. номз. дисс. – Тошкент, 1965); Рузикулова М. Идиоматика узбекского языка: автореф. дисс. канд. филол. наук. – Самарканд, 1966, – С. 22; Хаитметов К. Фразеологик бирликларда сўзлар тартиби масаласига доир // ЎТА, 1966. – №2. – Б.45-53.; Каххарова Х. Фразеология Абдуллы Кадыри: автореф. дисс. канд. филол. наук. – Ташкент, 1985, – С. 24; Йўлдошев Б. Ҳозирги ўзбек адабий тилида фразеологик бирликларнинг функционал-услубий хусусиятлари: филол. фанлари д-ри дисс. автореф. – Тошкент, 1993. – 50 б.; Усмонова Ш. Фразеологик бирликларда маъно ва қурилиш номуносивлиги // ЎТА, 1998. – №1. – Б.77.

<sup>39</sup> Рахматуллаев Ш. Некоторые вопросы узбекской фразеологии: Автореф. ди сс.... д-ра. филол. наук. – Ташкент, 1966. – 47 с.; Рахматуллаев Ш. Основные грамматические особенности образных глагольных фразеологических единиц современного узбекского языка: Автореф. дисс... канд. филол. наук. – М., 1952. – 19 с.; Рахматуллаев Ш. Ўзбек тилида феъл фраземаларининг боғлашуви. – Тошкент: Университет, 1992. – Б. 126.

<sup>40</sup> Маматов А.Э. Ўзбек тили фразеологизмларининг шаклланиши масалалари: filologiya fanlari doktori dissertatsiyasi avtoreferati. – Тошкент, 2000. – 254 б.

<sup>41</sup> Жуманазарова, Г. Фраземаларнинг нутқда куллаирилиши / Г.Жуманазарова [ва бошқ]. – Тошкент: Turon zamin ziyo. 2017. – 128 б.

<sup>42</sup> Йўлдошев Б. Ҳозирги ўзбек адабий тилида фразеологик бирликларнинг функционал-услубий хусусиятлари. Филол. фан. д-ри. дисс. – Тошкент, 1993.; Йўлдошев М. Чўлпоннинг халқ ибораларидан фойдаланиш маҳорати // Ўзбек тили ва адабиёти. 1999. 6-сон.; Маматов А., Очилов Д. Фраземаларнинг синтактик ва деривацион моделлари // Стилистика ва прагматика, 2-кисм. – Самарканд: СамДУ, 2010. – Б. 56-57.; Маматов А. Фразеологик стилистика масалари. – Тошкент, 1991.; Маматов А. Фразеологизмларнинг шаклланиш асослари. – Тошкент: Навруз, 1996.; Рахматуллаев Ш. Некоторые вопросы узбекской фразеологии: Автореф. ди сс.... д-ра. филол. наук. – Ташкент, 1966. – 47 с.; Рахматуллаев Ш. Основные грамматические особенности образных глагольных фразеологических единиц современного узбекского языка: Автореф. дисс... канд. филол. наук. – М., 1952. – 19 с.; Рахматуллаев Ш. Ўзбек тилида феъл фраземаларининг боғлашуви. – Тошкент: Университет, 1992. – Б. 126.

<sup>43</sup> Алматова Ш. Ўзбек тили фраземаларининг компонент тахлили: филол. фан. ном. ... дис. автореф. – Тошкент: ЎЗР ФА ТАИ, 2008. – 22 б.; Вафоева М. Ўзбек тилида фразеологик синонимлар ва уларнинг структурал-семантик тахлили: филолог. фан. номз. диссер. автореф. – Тошкент, 2009.; Насиров А. Француз, ўзбек ва рус тилларидаги проverbs фразеологизмларнинг семантик-стилистик ва миллий-маданий хусусиятлари: филол.

Лингвистические основы создания авторского корпуса узбекского языка были изучены Ш. Хамроевой; база синонимичных слов Национального корпуса узбекского языка А. Эшмуминовым; проблематика отражения фразеологических единиц узбекского языка в корпусе и формирование лингвистической базы данных Г. Бегматовой; лингвистические принципы построения учебного корпуса узбекского языка У. Холияровым; принципы создания корпуса узбекского языка Н. Абдурахмоновой и Г. Тоировой; теоретические и прикладные основы формирования корпуса интернет-текстов на узбекском языке О. Абдуллаевой; использование компьютерных методов при создании Национального корпуса узбекского языка А. Рахмановой; принципы и лингвистическое обеспечение создания стилистических тегов для единиц узбекского корпуса Д. Эловой; создание базы семантических тегов для авторского корпуса произведений Алишера Навои Н. Гуломовой; аннотирование грамматических форм времени в корпусе узбекского языка Э. Хонназаровым; проблемы токенизации, стемминга и лемматизации языковых единиц узбекского языка З. Хусаиновой; а также лингвистические основы и модели лексико-семантической разметки номинативных единиц в корпусе узбекского языка Д. Ахмедовой. Тем не менее, лингвистическое и программное обеспечение разметки фразеологических единиц узбекского языка до настоящего времени не стало предметом ни одного монографического исследования.

**Связь темы исследования с планами научно-исследовательской работы высшего учебного заведения, в котором выполнена диссертация.** Диссертация выполнена в соответствии с научно-исследовательским планом Ташкентского государственного университета узбекского языка и литературы по теме “Автоматическая обработка единиц узбекского языка, разработка корпуса узбекского языка и перспективы морфологической разметки”.

**Цель исследования** заключается в разработке лингвистических и программных основ тегирования фразеологических единиц узбекского языка в корпусе.

**Задачи исследования:**

изучение опыта тегирования фразеологизмов в мировой компьютерной лингвистике; обобщение результатов исследования лингвистических особенностей и речевой ценности фразеологизмов в узбекском языкознании; обоснование значимости тегирования фразеологизмов в языковом корпусе;

разработка алгоритмов токенизации и лемматизации фразеологических единиц узбекского языка, формирования базы данных; определение лингвистических подходов к морфологической разметке фразеологических единиц узбекского языка;

применение статистических и основанных на правилах методов в алгоритме лемматизации фразеологизмов;

определение алгоритма применения базы данных и статистического метода при морфологической разметке фразеологических единиц; предложение техники выявления и устранения омонимии во фразеологизмах.

**Объектом исследования** являются фразеологические единицы, представленные в словарях фразеологических единиц узбекского языка.

Предметом исследования является разработка лингвистических основ и программного обеспечения для тегирования фразеологических единиц узбекского языка.

**Методы исследования.** В исследовании использовались методы статистического анализа, компонентного анализа, описания и классификации.

**Научная новизна исследования** заключается в следующем:

в результате исследований, проведённых на основе языковых корпусов, были проанализированы эксперименты по тегированию фразеологизмов, выявлены специфические лингвистические особенности фразеологических единиц узбекского языка; изучены их составные компоненты, связь с частями речи, устойчивые и вариативные формы, контекстуальные проявления, а также возможности их автоматического распознавания;

актуальность тегирования фразеологических единиц в составе языкового корпуса обоснована с практической точки зрения; определены лингвистические подходы к морфологической разметке узбекских фразеологических единиц; разработан специальный алгоритм для лемматизации узбекских фразеологизмов. В связи с формированием базы данных фразеологических единиц расширены возможности автоматической обработки фразем и их интеграции в лингвистические ресурсы;

разработаны статистические и основанные на правилах методы лемматизации фразеологических единиц, создана специализированная система морфологических тегов для их идентификации и структурного анализа; эффективность статистического подхода на основе моделей n-грамм для распознавания фразеологических единиц обоснована в практическом аспекте;

в корпусе узбекского языка разработаны и реализованы алгоритмы и правила для выявления грамматических вариантов фразеологических единиц с использованием метода n-грамм, а также для выявления и устранения омонимии в составе фразем.

**Практические результаты исследования** заключаются в следующем:

разработана система тегов для фразеологических единиц узбекского языка;

проведено тегирование фразем, имеющих в корпусе узбекского языка;

сформирована база данных фразеологических единиц узбекского языка.

**Достоверность результатов исследования** объясняется применёнными подходами, моделями и алгоритмами, использованием методов и теоретических данных, заимствованных из надёжных научных источников, обоснованностью проведённых анализов с помощью рациональных и

комплексных методов, внедрением теоретических положений и выводов в практику, а также подтверждением полученных результатов уполномоченными структурами.

#### **Научная и практическая значимость результатов исследования.**

Научная значимость диссертации определяется разработанным алгоритмом выявления фразем - единиц, состоящих из нескольких токенов, выражающих одно лексическое значение, что позволяет автоматически распознавать и тегировать фразеологические единицы узбекского языка в корпусе. Это проявляется в подходах к идентификации языковых единиц в корпусе, включая фраземы, их лемматизацию, POS-тегирование и распознавание как целостных единиц в тексте.

Практическая значимость исследования заключается в том, что оно представляет собой поэтапное описание процесса автоматического распознавания, лемматизации и тегирования фразеологических единиц в корпусе узбекского языка; может служить базой данных при разработке общих принципов автоматического тегирования единиц корпуса узбекского языка; выполняет функцию теоретического материала для новых исследований в области корпусной лингвистики; а полученные научные результаты могут быть использованы в разработке учебных программ и пособий по дисциплинам направления компьютерной лингвистики, а также в создании учебных пособий по лексикологии.

**Внедрение результатов исследования: на основе научных результатов,** полученных в процессе тегирования системы фразем узбекского языка:

на основании проведённых исследований, выполненных с использованием языковых корпусов, были проанализированы эксперименты по тегированию фразеологизмов, выявлены специфические лингвистические характеристики фразеологических единиц узбекского языка, их структурные компоненты, связи с частями речи, устойчивые и изменяемые формы, контекстуальные проявления, а также возможности автоматического распознавания, и данные научные выводы были внедрены в деятельность Государственного учреждения «Центр повышения квалификации и обучения основам делопроизводства на государственном языке» при Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои (согласно протоколу Центра от 4 февраля 2025 года). В результате, при содействии Центра внедрения электронного обучения при учебных заведениях была сформирована база данных, содержащая 1800 тегированных фразеологических единиц узбекского языка с использованием имеющихся электронных ресурсов.

актуальность тегирования фразеологизмов в языковом корпусе была обоснована с прикладной точки зрения; научные выводы, касающиеся лингвистических подходов к морфологическому тегированию фразеологических единиц узбекского языка, алгоритма, направленного на лемматизацию фразеологизмов, и базы данных фразеологических единиц, а

также возможностей их автоматической обработки и интеграции в лингвистические ресурсы, были использованы в научном проекте по теме «UzUDT: Универсальный корпус дерева зависимостей для узбекского языка и его семантический анализ», реализованном в 2022–2024 годах в Национальном университете Узбекистана имени Мирзо Улугбека по гранту REP-25112021/113 (согласно справке №04/11-5273 от 23 апреля 2025 года, выданной университетом). В результате была создана возможность корректного распознавания фразеологических единиц в составе предложения на основе разработанных математических моделей, алгоритмов и программного обеспечения для лемматизации фразеологизмов узбекского языка, а также достигнута точность синтаксического анализа 1800 теггированных фразеологических единиц с применением методов и алгоритмов семантического конкорданса на уровне 92 %.

научные выводы, касающиеся статистических и правил-основанных методов лемматизации фразеологизмов, системы морфологических тегов для идентификации и структурного анализа фразеологических единиц, а также статистического подхода, основанного на n-граммных моделях, были использованы в прикладном проекте «Создание образовательного корпуса узбекского языка», реализованном в 2021–2023 годах в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои по гранту IZ-2021020813 (согласно справке №04/1-4250 от 24 декабря 2024 года, выданной университетом). В результате, алгоритм лемматизации фразеологизмов узбекского языка и база данных фразеологических единиц были применены для формирования поисковой системы образовательного корпуса узбекского языка, а также обеспечено эффективное использование алгоритма для выявления и устранения омонимии в составе фразеологизмов.

алгоритм выявления грамматических вариантов фразеологических единиц методом n-грамм, алгоритм распознавания омонимии в фразеологизмах и соответствующие правила были использованы в прикладном проекте «Создание образовательного корпуса узбекского языка», реализованном в 2021–2023 годах в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои по гранту IZ-2021020813 (согласно справке №04/1-4250 от 24 декабря 2024 года, выданной университетом). В результате, функциональность n-граммного поиска в системе образовательного корпуса узбекского языка была усовершенствована.

**Апробация результатов исследования.** Результаты диссертации были обсуждены на 11 научно-практических конференциях, включая 8 международных и 3 республиканскую научно-практическую конференцию.

**Публикация результатов исследования.** По теме диссертации опубликованы всего 26 работы, включая 13 статей в научных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан

для публикации основных научных результатов докторских диссертаций, из них 10 - в республиканских и 3 - в зарубежных научных журналах.

**Структура и объём исследования.** Диссертация состоит из введения, трёх основных глав, общей заключительной части и списка использованной литературы; общий объём составляет 124 страницы.

## ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении диссертации обоснованы актуальность и необходимость выбранной темы, определены цели и задачи исследования, а также описаны объект и предмет исследования. Отражено соответствие работы приоритетным направлениям развития науки и технологий в Республике Узбекистан, изложены научная новизна и практические результаты исследования, раскрыто научно-практическое значение полученных результатов. Также приведены сведения о внедрении результатов исследования в практику, опубликованных работах, а также о структуре и объёме диссертации.

**Первая глава исследования** носит название **“Общая характеристика фразеологических единиц узбекского языка”**. В первом параграфе главы рассматривается *взаимосвязь мировой фразеологии и корпусной лингвистики*.

Во мировой фразеологии фразеологические единицы различных языков исследовались с разных аспектов. Фразеологические исследования служат теоретической базой различного характера для корпусной лингвистики. С одной стороны, результаты таких исследований служат теоретической основой для теггирования (аннотирования) фразеологических единиц в языковых корпусах; с другой стороны, эти исследования имеют значение для компьютерной лексикографии, так как на их основе подготавливается база для создания электронных фразеологических словарей различных языков.

Взаимосвязь лингвистических корпусов и фразеологии отчётливо проявляется в трёх основных аспектах:

во-первых, корпусные исследования в области фразеологии;  
во-вторых, отражение фразеологических единиц в языковых корпусах;  
в-третьих, использование корпусов как лингводидактического инструмента при изучении фразеологии.

В научной литературе выделяются следующие основные виды работы с корпусами:

–исследования, ориентированные на корпус (corpus-driven);  
–исследования, основанные на корпусе (corpus-based);  
–исследования, использующие корпус в качестве иллюстративного материала (corpus-illustrated, corpus-informed)<sup>44</sup>.

Хотя в ряде исследований границы между этими типами подходов размываются, гетерогенный характер корпусного подхода даёт хорошие

---

<sup>44</sup> Добровольский Д.О. Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов. Вестник Санкт-Петербургского университета. Язык и литература, (2020). 17(3), 398–411. <https://doi.org/10.21638/spbu09.2020.303>

результаты, поскольку при анализе с использованием корпуса каждый случай рассматривается индивидуально. Таким образом, если исследователь использует корпус в качестве источника иллюстративного материала (corpus-illustrated подход), его основная задача заключается, прежде всего, в выборе контекста парадигматических примеров. Этот принцип особенно актуален в лексикографических процессах. Словарь, составленный на основе иллюстративного материала, извлечённого из корпуса, отличается относительно большей объективностью по сравнению со словарём, основанным на языковой интуиции или случайно найденных в архиве лексикографа иллюстративных примерах. По мнению Д.О.Добровольского, принцип использования корпуса для работы с иллюстративным материалом не означает, что лексикограф освобождается от необходимости принимать индивидуальные решения в каждом конкретном случае. Напротив, преимущество этого принципа заключается в том, что он расширяет возможности выбора и позволяет, опираясь на объективные данные, определить степень употребления конкретной формы<sup>45</sup>.

Ряд исследований, посвящённых фразеологии в мировой корпусной лингвистике, появился на основе анализа результатов, полученных с использованием корпуса или на базе корпуса. В подобных исследованиях языковые корпуса и представленные в них фразеологические единицы выполняют роль либо лингвистического материала, либо метода исследования (метода корпусного анализа). При этом основное внимание уделяется анализу фразеологических единиц, а языковые корпуса выступают лишь в качестве инструмента исследования.

Существует ещё одно направление исследований, связанных с лингвистическими корпусами и фразеологией, которое посвящено теггированию фразеологических единиц в языковых корпусах, их отображению в корпусах, разработке систем поиска, а также использованию языковых корпусов для поиска и анализа фразем. В таких исследованиях на первый план выходят вопросы создания корпусов, их использования и отражения фразеологических единиц в лингвистических корпусах.

В мировой лингвистике широко применяются методы исследования фразеологических единиц, основанные на корпусах. Данный метод ценен тем, что анализ проводится обоснованно, а материалы анализа - корпусные данные - отличаются объективностью. Такой подход обеспечивает высокую точность получаемых результатов. Анализ многочисленных публикаций свидетельствует о популярности корпусного анализа различных лексических и фразеологических единиц. Однако проблема теггирования фразеологических единиц в языковых корпусах остаётся слабо изученной областью. Аннотирование фразеологических единиц в лингвистических корпусах существенно расширяет возможности проведения корпусных исследований.

Во втором параграфе главы анализируется изучение *фразеологизмов в узбекском языкознании, их лингвистические особенности и речевая ценность*.

---

<sup>45</sup> Shu muallif, ko'rsatilgan manba. – B. 399-400.

Фразеологические единицы являются неотъемлемой частью лексики языка, помогают выражать определённые значения, создавать образы и повышать экспрессивность речи. Фразеологизмы также обладают лингвокультурной спецификой, и в других языках их эквиваленты могут отсутствовать. Фразеологические единицы конкретного языка формируются и развиваются в процессе языкового общения и культурного прогресса общества. К факторам, влияющим на формирование и развитие фразеологических единиц, можно отнести: (1) исторический контекст, (2) культурные особенности, (3) языковые процессы, (4) метафору и образы, (5) влияние других языков. Формирование и развитие фразеологических единиц языка представляют собой динамический процесс, отражающий языковую и культурную эволюцию общества. Они являются важной частью языка и способствуют обогащению и выразительности речи.

При изучении фразеологизмов узбекского языка важную роль играют различные составленные словари и учебные пособия. Особенности фразеологических единиц узбекского языка в разной степени освещены в научных статьях.

Экспрессивные свойства фразеологизмов, вопросы их перевода, окказиональные фразеологизмы в языке прессы, проблемы фразеологии в толковом словаре узбекского языка, семантический анализ личностно-характеризующих фразем, тематические группы фразем разговорной речи, соотношение между фразеолектами и лексикографией, использование фразеологизмов в художественной речи, сходные глагольные фразеологические единицы в узбекском и афганском языках, исследования, посвящённые отношению фразеологизмов к нормам литературного языка, - всё это служит важным материалом для выявления различных характеристик фразеологических единиц, их теггирования и включения в базы данных узбекских языковых корпусов.

В третьем параграфе первой главы анализируются *лингвистические основы теггирования фразеологических единиц в корпусе языка и вопросы, связанные с базой данных*. Для обеспечения фразеологических единиц в лингвистической базе данных языковых корпусов достаточной информацией целесообразно определить их лексико-семантические, морфологические, грамматические, стилистические и другие характеристики. Фразеологические единицы узбекского языка с этой точки зрения были всесторонне изучены. Имеющийся научный материал - систематизация данных - является достаточной основой для нашей работы по теггированию фразеологических единиц в корпусе. С грамматической точки зрения одни фразеологические единицы эквивалентны одному слову и выражают единое значение, другие соотносятся со словосочетанием, а третьи - с предложением, выражая относительно завершённую мысль. Исходя из этих характеристик, фразеологические единицы первоначально делятся на три группы (см. схему 1 в приложении). Соответственно, в ряд тегов вводятся теги, отражающие

грамматические признаки фразеологических единиц: слово-оборот, сочетание-оборот и предложение-оборот.

Фразеологизмы могут выступать в предложении в роли любых членов:

а) подлежащее: *Otaxonning yuziga oyoq qo'yish bizning odatimizda yo'q.* (А.Мухтор.);

б) сказуемое: *Shunaqa vaqtda odamning yuragi qinidan chiqib ketadi* (qo'rqadi ma'nosida) (А.Каххар.);

в) определение: *O'rinbosarning bunday xushomadlari oldida og'zining tanobi qochadigan odati bor edi* (в значении: начинал улыбаться) (из газеты.);

г) дополнение: *Qo'rqar edi sho'rlik ko'zni yumishdan* (в значении: умереть) (Х.Олимджон.);

д) обстоятельство: *Hozir dunyoni boshiga ko'tarib buvim kelar* (в значении: устроит скандал, поднять шум).

Фразеологические единицы по синтаксическому строению встречаются в виде словосочетаний, простых и сложных предложений (см. схему 1). Определение их типа по структуре важно для процессов лемматизации и теггирования.

**Вторая глава исследования** носит название **“Статистический анализ фразем в корпусе узбекского языка”**. Первый параграф посвящён изучению морфологических форм и вариантов узбекских фразем с позиций традиционного и корпусного анализа.

Вариантность - явление, присущее всем языковым единицам, включая фразеологические единицы. Основную часть многовариантных фразем составляют двух- и трёхвариантные выражения. В целом, вариантность во фразеологических единицах выражена сильнее, чем в других языковых единицах: количество вариантов одной фраземы может достигать десяти. В результате вариантности одна фразема не превращается в другую - сохраняются единство образа и лексическое значение, характерные для данной фраземы. Вариантность реализуется в пределах целостности, присущей фраземе, и только в той мере, в какой это допускает данная целостность. В противном случае фразема либо исчезает, либо образуется новая фразема.

Во фразеологических вариантах различают лексические и грамматические варианты. Вариантом, образованным путём изменения компонента, выраженного словом самостоятельной части речи, считается *лексический вариант*. При этом компонент может заменяться, добавляться или опускаться. Наиболее сложной является вариантность, основанная на лексической замене<sup>46</sup>.

Вариантом, образующимся путём грамматического изменения состава фраземы, считается *грамматический вариант*. Грамматические изменения, связанные с внешним грамматическим строением (с включением фраземы в контекст), не рассматриваются как вариантность. В частности, изменения существительных в составе фраземы по числу и падежу, изменения глагольных фразем по спряжению, с добавлением показателей принадлежности, наклонения или времени относятся не к вариантности, а к

<sup>46</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

*парадигматическим формам*<sup>47</sup>. Такие варианты, точнее, парадигматические формы, в языке теоретически бесконечны. Если искать любую фразу в языковых корпусах, можно встретить сотни её грамматических или морфологических форм. Например, при поиске фраземы “*achchig‘i kelmoq*” в корпусе узбекского языка была выявлена следующая парадигма:

<i>achchig‘i keldi</i>	<i>achchig‘i kelgan</i>
<i>achchig‘i kelib</i>	<i>achchig‘i kelyapti</i>
<i>achchig‘im keladi</i>	<i>achchig‘i keladi-da</i>
<i>achchig‘ini keltiradigan</i>	<i>achchig‘i keladi</i>
<i>achchig‘i kelur</i>	<i>achchig‘im kelgan</i>
<i>achchiqlangan kelinni</i>	<i>achchig‘imni keltirdi</i>
<i>achchig‘ini keltirdi</i>	<i>achchig‘lari kelgan-da</i>
<i>achchig‘ing keladi</i>	<i>achchig‘i kelgandek</i>
<i>achchig‘i kelsa</i>	

Если фразеологическая единица в указанных выше типах парадигматических форм имеет определённые ограничения, тогда грамматическая вариантность считается уместной. Например, употребление глагольной фраземы в утвердительной или отрицательной форме (*ko ‘kka sovurdi, ko ‘kka sovurmadi*) является общим случаем и не рассматривается как грамматическая вариантность. В то же время существуют фраземы, которые функционируют преимущественно или исключительно в одном аспекте. Например, выражение *alifni tayoq deya olmaslik* обычно употребляется в отрицательной форме, но иногда встречается и в утвердительной: *alifni tayoq deydigan*. В этом случае уже необходимо признать наличие фразеологического варианта. Грамматические варианты в прямом смысле возникают в результате изменений во внутреннем грамматическом строении фраземы (в частях, не связанных с контекстом). Например, образование залоговых (видовых) форм в глагольных фраземах и связанные с этим грамматические изменения - это один из типов такой вариантности. Так, в глагольной фраземе *adab(i) bermoq* глагольный компонент выражен в определённой степени, а компонент в винительном падеже образует конструкцию с дополнением. Та же фразема может использоваться в форме *adab(i) berildi*. Изменение залога резко меняет внутреннее синтаксическое строение фраземы: конструкция, эквивалентная словосочетанию, превращается в конструкцию, эквивалентную предложению, что обуславливает последующие грамматико-семантические трансформации<sup>48</sup>.

Грамматическая вариантность, так же как и лексическая, в основном возникает за счёт замены, опущения или добавления какого-либо грамматического элемента в составе фраземы. Особым явлением в

<sup>47</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

<sup>48</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

грамматической вариантности является вариантность, возникающая на основе порядка расположения слов-компонентов:

**4) вариантность, возникающая на основе замены:** *o'tirishga joy topolmaslik // o'tirgani joy topolmaslik* - в данном случае происходит замена отглагольного существительного, образованного с помощью суффикса *-gani* (цель действия), на форму с суффиксом *-sh* (действие как процесс) и направительным падежом (*-ga*); *qo'yn(i)ni puch yong'oqqa to'ldirmoq // qo'ynini puch yong'oq bilan to'ldirmoq* - происходит замена падежной формы (*-ga*) на конструкцию с послелогом *bilan*.

**5) вариантность, возникающая на основе добавления:** *holdan toymoq // holdan toydirmoq* - во втором случае добавлен каузативный аффикс, указывающий на причинность действия; *ko'ngl(i) bo'shadi // ko'ngl(i) bo'shashdi* - к глаголу добавлен аффикс, выражающий ослабленное, менее интенсивное проявление действия.

**6) вариантность, возникающая на основе опущения:** *ko'z(i) ilindi, ko'z ilindi* в первой конструкции присутствует форма принадлежности, которая во втором варианте опущена, что не влияет существенно на значение, но формирует морфологическую вариацию<sup>49</sup>.

Как видно из приведённой информации, для выявления вариантов каждой фраземы и их включения в словарь необходимо отслеживать их речевое употребление на основе большого датасета, точнее - языкового корпуса. Осуществить это вручную невозможно. В корпусах, предназначенных для работы с большими данными, теггирование фразеологических единиц позволяет выявлять каждую грамматическую или морфологическую форму фразем.

Сравним следующие определения явления фразеологической вариантности:

1. “Фразеологические варианты - это различные формы инвариантной фразеологической единицы, отличающиеся по смысловому составу и содержанию, стилистическим и синтаксическим функциям, представляющие собой частичные лексические или морфологические изменения, а также изменения формы слова или порядка слов”<sup>50</sup>.

2. Структурным вариантом идиомы называют фразеологическую единицу, которая не имеет существенных различий в составе и структуре<sup>51</sup>.

3. Фразеологические варианты отличаются стилистической окраской и экспрессивностью, и в разных контекстах употребляются их наиболее подходящие формы<sup>52</sup>.

Обобщая приведённые определения, можно сказать, что фразеологические варианты представляют собой речевые реализации одного

<sup>49</sup> Рахматуллаев Ш. Ўзбек тилининг изохли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

<sup>50</sup> Кунин А.В. Основные понятия английской фразеологии как лингвистической дисциплины и создание англо-русского фразеологического словаря. Дисс. д-ра филол. наук. – М. 1964. – С. 442.

<sup>51</sup> Амосова Н.Н. Основы английской фразеологии. – Л., 1963. – С. 100.

<sup>52</sup> Авалиани Ю.Ю., Ройзензон Л.И. О разграничение синонимии и вариантности в области фразеологических единиц // Вопросы фразеологии и составления фразеологических словарей. – Баку, 1968. – С. 72.

инварианта, которые в зависимости от речевого стиля и экспрессивной окраски встречаются в текстах в разных вариантах.

Ю.П.Игнатович и Т.Н.Федуленкова, подчёркивая существование фразеологической вариантности в языке, выделяют следующие её виды в составе фразем:

1. Лексическая вариативность.
2. Лексико-морфологическая вариативность.
3. Квантитативная вариативность
4. Лексико-квантитативная вариативность.
5. Синтаксическая вариативность.
6. Лексико-синтаксическая вариативность.
7. Синтаксико-квантитативная вариативность.<sup>53</sup>

Кроме того, авторы в статье установили, что наибольший процент вариантности в английских фраземах приходится на лексическую вариативность.<sup>54</sup>

До настоящего времени в лингвистике мы не встречали исследований, основанных на корпусном подходе, направленных на выявление вариантов фразеологических единиц. На самом деле, если в языковых корпусах фраземы аннотированы или имеется функция n-грамм, можно определить грамматические и лексические варианты, а также грамматические парадигмы фразем; результаты такого анализа считаются более точными и надёжными. Ниже мы проанализируем результаты поиска в корпусе узбекского языка, а также морфологические формы и грамматические варианты некоторых выражений из фразеологического словаря<sup>55</sup> узбекского языка.

В фразеологическом словаре для выражения *abjag 'ini chiqarmoq* указана одна морфологическая форма: *abjag 'i chiqdi*<sup>56</sup>. Однако в результате поиска по корпусу узбекского языка были выявлены следующие 10 морфологических форм этого выражения:

<i>abjag 'i chiqqan</i>	<i>abjag 'imni chiqarib</i>
<i>abjag 'i chiqib</i>	<i>abjag 'ini chiqar</i>
<i>abjag 'ing chiqqan</i>	<i>abjag 'ini chiqarganga</i>
<i>abjag 'i chiqdi</i>	<i>abjag 'ini chiqargan</i>
<i>abjag 'ingni chiqaraman</i>	<i>abjag 'ini chiqaradi</i>

Из полученных результатов видно, что морфологические формы могут изменяться в обеих компонентах этого выражения. Часть “*abjag '(i)*” встречается в различных формах принадлежности и падежа, а форма “*chiqarmoq*” - во временных, причастных и деепричастных формах. Несмотря

<sup>53</sup> Игнатович Я.П., Федуленкова Т.Н. Вариант фразеологической единицы как показатель ее тождества (на материале фразеологии с компонентом – глаголом отчуждения) // [https://journals.tsu.ru/uploads/import/1583/files/37\\_035.pdf](https://journals.tsu.ru/uploads/import/1583/files/37_035.pdf)

<sup>54</sup> Игнатович Я.П., Федуленкова Т.Н. Вариант фразеологической единицы как показатель ее тождества (на материале фразеологии с компонентом – глаголом отчуждения) // [https://journals.tsu.ru/uploads/import/1583/files/37\\_035.pdf](https://journals.tsu.ru/uploads/import/1583/files/37_035.pdf)

<sup>55</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – 407 б.

<sup>56</sup> Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 22.

на встречающиеся различные морфологические формы, это выражение не образует грамматического варианта.

Выражение “*andaza olmoq*” в фразеологическом словаре не приведено. Во введении к словарю указано, что в нём представлены только высокочастотные выражения или выражения, имеющие синонимические, антонимические пары и варианты<sup>57</sup>. Однако, как показывают результаты поиска по корпусу, выражение “*andaza olmoq*” в узбекском языке нельзя отнести к низкочастотным. В корпусе были выявлены следующие морфологические формы этого выражения:

<i>andaza olish</i>	<i>andaza olgin</i>
<i>andaza olgan</i>	<i>andaza oladi</i>
<i>andaza olsa</i>	<i>andaza oladigan</i>
<i>andaza olgandek</i>	<i>andaza olinayotganidan</i>
<i>andaza olingan</i>	

На этой основе целесообразно включить данное выражение также в фразеологический словарь.

#### *Определение частотности фразем с помощью корпуса узбекского языка*

В корпусе невозможно выявить идиомы или фраземы с помощью простого поиска. При поиске фразем методом n-грамм в список результатов часто попадают также словосочетания, которые не являются фразеологическими единицами в данном языке. Хотя метод n-грамм считается наиболее оптимальным для поиска фразем, тем не менее, обязательно проводится ручной анализ результатов и очистка списка (удаление единиц, не являющихся фразеологизмами). Например, поиск по выражению “*baxti ochilmoq*” дал следующий результат:

- |                                |                             |
|--------------------------------|-----------------------------|
| 1) <i>baxtim ochilmadi</i>     | 6) <i>baxting ochilsin</i>  |
| 2) <i>baxti ochilishi</i>      | 7) <i>baxtiqaro oqsochi</i> |
| 3) <i>baxting ochilmadi-da</i> | 8) <i>baxtidan ochish</i>   |
| 4) <i>baxti ochilmayapti</i>   | 9) <i>baxti ochilsa-ku</i>  |
| 5) <i>baxti ochilib</i>        |                             |

Как видно из приведённого списка, данное выражение встречается в корпусе 9 раз, или же слова “*baxt*” и “*ochilmoq*” появляются столько раз. Однако выражение “*baxtiqaro oqsochi*” не относится к данной парадигме. Следовательно, после удаления ошибочного словосочетания из результатов поиска методом n-грамм в корпусе узбекского языка, выражение “*baxti ochilmoq*” имеет частотность 8.

Фразеологический слой узбекского языка чрезвычайно богат, и его элементы используются с разной степенью частотности: одни фраземы активно употребляются, а частотность других невысока. Для определения частотности узбекских фразем с помощью корпуса узбекского языка<sup>58</sup> требуется, чтобы фраземы были предварительно размечены тегами частей речи (POS-теги). Для проведения такого поиска была сначала сформирована

<sup>57</sup> Рахматуллаев Ш. Ўзбек тилининг изохли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – Б. 13-15.

<sup>58</sup> <https://uzschoolcorpara.uz/>

база данных узбекских фразем, фраземы были приведены к словарной форме, которая была признана лемматизированной формой выражения, а затем они были размечены как существительное-оборот, прилагательное-оборот, глагольное-оборот (об этом подробнее будет написано в следующей главе).

Если фраземы в корпусе лемматизированы, становится возможным осуществлять их поиск в корпусе. В этом случае можно определить как низкочастотные, так и высокочастотные единицы. В таблице 1 ниже представлена частотность некоторых фразем узбекского языка по результатам поиска в корпусе (см. таблицу 1).

№	Фразема	классификация фразем по значению	частотность
1	abjag‘i chiqqan	глагольная фразема	10
2	aksiga olib	глагольная фразема	2
3	andaza olish	глагольная фразема	9
4	aql bovar qilmaydigan	прилагательная фразема	26
5	bahri-dilim ochildi	глагольная фразема	17
6	barmog‘ini tishlamoq	глагольная фразема	

**Таблица 1. Примеры частотности употребления узбекских фразем по данным корпуса [uzschoolcorpora.uz](http://uzschoolcorpora.uz)**

Из таблицы 1 видно, что частотность употребления фразем в узбекском языке неодинакова. Среди пяти отобранных фразем относительно высокой частотностью отличаются выражения, передающие признак, такие как “*aql bovar qilmaydigan*” (26 вхождений) и “*bahri-dili ochilmoq*”, а низкая частотность зафиксирована у фраземы “*aksiga olib*”. Эта вывод, безусловно, относителен, поскольку результаты были получены из учебного корпуса узбекского языка, который включает в себя, главным образом, тексты школьных учебников, художественные произведения, предназначенные для школьного возраста, а также материалы с интернет-сайтов. В подобных текстах низкая встречаемость фразем является обычным явлением.

Третья глава исследования называется “**Программные основы теггирования фразеологических единиц узбекского языка**”. В первом параграфе рассматривается вопрос *формирования базы данных фразеологических единиц на основе корпуса узбекского языка*.

Лингвистическая база данных - это упорядоченный набор данных, содержащий слова, выражения, грамматические формы, идиомы и другие языковые элементы. Такие базы данных могут использоваться для различных целей, в том числе:

1. В лексикографии: для создания словарей и справочников;
2. В обработке естественного языка: для поддержки технологий, связанных с автоматическим переводом, распознаванием речи и генерацией текста;

3. В лингвистических исследованиях: для анализа структуры языка, изменений и использования языковых единиц;

4. В образовании: для создания учебных материалов и инструментов по изучению языков.

База данных может состоять из таблиц с определёнными полями, включающих описание каждой языковой единицы, её роль в предложении, иллюстративные примеры и другие метаданные.

Lingvistik ma'lumotlar (til birliklari)ning ma'lumotlar bazasini yaratish bir necha bosqichda amalga oshiriladi. Bu kabi ma'lumotlar bazasini yaratishning bosqichma-bosqich qadamlarini quyidagicha izohlaymiz:

**1. Определение целей и требований.** Перед созданием базы данных необходимо определить, для чего она будет использоваться:

- какие языковые единицы (слова, выражения, грамматические конструкции) будут включены в базу данных;
- какая информация будет храниться о каждой единице (определение, часть речи, примеры, синонимы, антонимы и т. д.);
- какие критерии поиска и фильтрации данных будут использоваться.

**2. Выбор платформы.** Для хранения базы данных выбирается и определяется платформа:

- реляционные базы данных (например, MySQL, PostgreSQL, SQLite);
- NoSQL базы данных (например, MongoDB);
- редакторы электронных таблиц для небольших проектов (например, Microsoft Excel, Google Sheets).

**3. Проектирование структуры базы данных.** Создаётся структура таблиц для базы данных. Основные таблицы могут включать:

- таблицу языковых единиц, содержащую основную информацию о каждой единице;
- таблицу значений (значения/толкования), в которой описываются определения, пояснения и примеры употребления;
- таблицу грамматических характеристик, содержащую информацию о частях речи, лице, числе и других грамматических признаках.

**4. Ввод данных.** Процесс сбора данных и внесения их в таблицы может выполняться вручную или быть автоматизирован с помощью скриптов (например, с использованием Python для анализа текстовых данных).

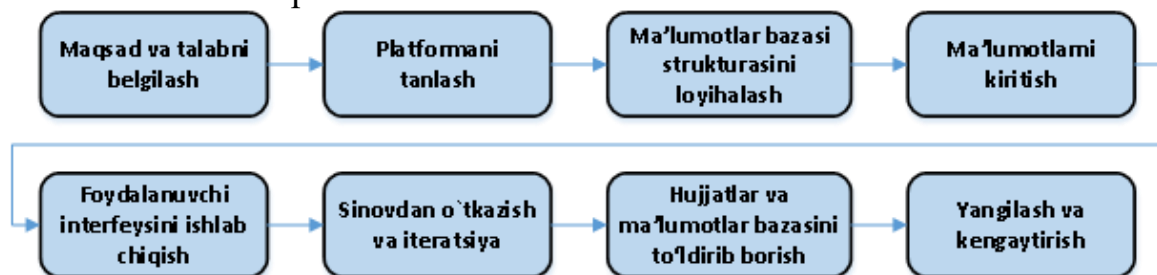
**5. Разработка интерфейса доступа.** Если база данных будет использоваться другими пользователями, целесообразно создать удобный интерфейс. Такой интерфейс можно реализовать в следующих формах:

- веб-приложение с формами для ввода и поиска данных;
- мобильное приложение.

**6. Тестирование и итерация.** Для обеспечения корректной работы базы данных необходимо провести её тестирование. Ошибки исправляются на основе отзывов пользователей.

**7. Документация и пополнение базы данных.** В процессе использования базы данных она обновляется за счёт добавления документации, новых языковых единиц и внесения исправлений.

**8. Обновление и расширение.** База данных регулярно обновляется путём добавления новых единиц, исправления ошибок и улучшения структуры в соответствии с потребностями пользователей.



**Схема 2. Этапы создания базы данных**

Выполнив указанные выше этапы, можно создать функциональную и эффективную лингвистическую базу данных.

Выше была представлена общая характеристика создания лингвистической базы данных. Ниже мы рассмотрим особенности построения фразеологической базы данных.

Фразеологическая база данных служит для обогащения лексикографических ресурсов, преподавания языка и проведения лингвистических исследований. При создании этого типа лингвистической базы данных необходимо учитывать ряд специфических особенностей.

### **1. Требования к содержанию базы данных:**

1. Определение типа фразеологической единицы (выражение, пословица, устойчивое словосочетание).

2. Определение атрибутов:

- фразеологизм (выражение);
- толкование;
- перевод (если база данных охватывает данные и на других языках);
- примеры;
- тематическая область, категория (например, природа, питание и др.);
- степень употребимости;
- синонимы и антонимы (если есть).

3. Требуется проектирование структуры базы данных:

Модель данных - формирование структуры базы данных. Например:

- таблица “выражения” (ID, фразеологизм, толкование, категория);
- таблица “примеры” (ID, ID фразеологизма, пример);
- таблица “переводы” (ID, ID фразеологизма, перевод, язык).

Создание базы данных фразеологических единиц с использованием лингвистического корпуса является эффективным способом сбора, анализа и построения базы данных фразеологизмов на основе реальных примеров, извлечённых из языка. Этот процесс реализуется на основе следующего алгоритма:

**1. Выбор лингвистического корпуса.** На этом этапе выбирается корпус, который будет служить источником данных для базы данных. Целесообразно выбирать лингвистический корпус, включающий различные текстовые материалы (например, новости, художественную литературу, научные статьи). Для удобного поиска и извлечения фразеологических единиц (идиом) можно использовать специализированное программное обеспечение для работы с корпусами, например Sketch Engine или AntConc.

**2. Извлечение фразеологических единиц.** Для поиска фразеологических единиц (устойчивых выражений) необходимо использовать функции поиска и фильтрации в выбранном корпусе (например, поиск сочетаний по шаблону “глагол + существительное”).

### **3. Автоматизация поиска.**

- Скрипты и алгоритмы: если объектом является большой объём текста, для автоматического извлечения выражений можно использовать скрипты на Python или других языках программирования с применением простых выражений или других методов обработки текста;

- Машинное обучение: для классификации и извлечения идиом из больших текстовых массивов эффективно использовать методы машинного обучения.

Это общее описание алгоритма и подходов к созданию базы данных фразеологических единиц. В данном параграфе будет подробно рассмотрен конкретный алгоритм разработки базы данных фразеологических единиц на основе корпуса узбекского языка.

Как отмечалось в предыдущих параграфах, объектом исследования является корпус узбекского языка, и фразеологическая база данных была сформирована на основе этого корпуса. Для этого в результате последовательных запросов был определён список каждой фразеологической единицы с её грамматической парадигмой, и на этой основе была создана база данных.

Для идентификации фразеологических единиц в базе данных корпуса узбекского языка используется следующая ER-схема (см. схему 2).

В данном параграфе мы рассмотрим, какое место занимают фразеологические единицы с различной частотностью в базе данных, а также проанализируем структуру и состав базы данных и статистику по фразеологическим единицам узбекского языка.

```
SELECT dbo.Token.Id, dbo.Token.Value, dbo.Token.CorrectVersion,  
dbo.WordToken.LemmaValue, dbo.WordToken.LemmaSuffixAffsPart,  
dbo.WordToken.StemValue, dbo.WordToken.StemPrefixAffsPart,  
dbo.WordToken.StemSuffixAffsPart, dbo.WordToken.RootValue
```

```
FROM dbo.Token INNER JOIN
```

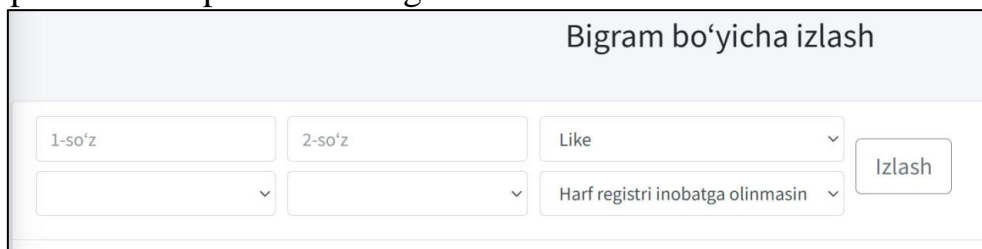
```
dbo.WordToken ON dbo.Token.Id = dbo.WordToken.TokenId
```

```
WHERE dbo.WordToken like % “ajali yet(moq)”
```

Например, для поиска выражения “ajali yetmoq” запрос выполняется по следующему алгоритму:

1. Необходимо открыть следующую ссылку:

<https://uznatcorpara.uz/uz/Bigram/Search>



**Рисунок 1. Окно поиска фразеологических единиц**

В данном интерфейсе необходимо ввести 2 токена (“ajali” и yet(moq)”) в соответствующие поля.

1. В разделе “**Harf registri inobatga olinsin/olinmasin**” следует выбрать соответствующий элемент.
2. В базе данных корпуса можно искать указанное словосочетание, используя параметры “**Like**” (%) или “**Regulyar ifoda**”.
3. После установки необходимых параметров поиска нужно нажать кнопку “**Izlash**”.
4. Результат будет представлен в следующем виде (см. схему 3 в приложении):

В результате была получена следующая грамматическая парадигма выражения “ajali yetmoq”

- |                    |                       |
|--------------------|-----------------------|
| 1. ajaling yetgani | 8. ajalimiz yetganida |
| 2. ajali yetgan    | 9. ajali yetgani      |
| 3. ajali yetib     | 10. ajali yetar       |
| 4. ajali yetgudek  | 11. ajalim yetguncha  |
| 5. ajalim yetib    | 12. ajaling yetmay    |
| 6. ajali yetmagan  | 13. ajali yetganini   |
| 7. ajali yetganda  |                       |

Как видно из списка, данные морфологические формы являются уникальными и не дублируются. В полученной парадигме отсутствуют ошибочные морфологические формы или варианты. Из этого следует, что в морфологическую базу данных будет включено 13 морфологических форм выражения “ajali yetmoq”.

На основе приведённого выше алгоритма была определена парадигма выражений *aksiga olmoq, anqoning urug'i, alamidan chiqmoq, aqli kirmoq* (см. таблицу 2 в приложении). Из таблицы 3 видно, что в корпусе, состоящем из 51890123 предложений по состоянию на 01.04.2024, частотность употребления всех фразеологических единиц различна: выражение *aksiga olmoq* встречается 2 раза, *anqoning urug'i* - 5 раз (один из которых оказался синонимом), *alamidan chiqmoq* - 17 раз, *aqli kirmoq* - 23 раза; список был автоматически сформирован.

Процесс формирования базы данных будет представлен в следующей последовательности в рамках бизнес-процесса (см. схему 4 в приложении).

Таким образом, в базу данных можно включить первоначальную (оригинальную) форму данной фразеологической единицы, её морфологическую парадигму, варианты и синонимы.

В корпусе, состоящем из 51890123 предложений по состоянию на 01.04.2024, было выявлено 23 коллокации выражения *aqli kirmoq*. Среди этих коллокаций встречаются такие, как *aqliy o 'tkirlik*, *aqlining o 'tkirligi*, которые не могут считаться фразеологическими единицами, то есть не являются вариантами данного выражения. Поэтому в процессе очистки эти коллокации будут удалены, и в базу данных будет внесено 21 морфологическое вариантное выражение *aqli kirmoq*. Поскольку вариант данного выражения уже указан в словаре и представляет собой лексический вариант (одно слово полностью отличается), также будет сформирована морфологическая парадигма его варианта.

В корпусном поиске морфологическая парадигма выражения *es(i) kirmoq* выглядит следующим образом: *esi kirganidan*, *esini kiritib*, *esingiz kirarli-chiqarli*, *esi kirdi-chiqdi*, *esini kiritish*, *esi kirib*, *esimiz kirmaydi*, *esi kirsas*, *esi kirmagan*, *esim kiradigan*, *esi kirgan*, *esing kirmadi*, *esingni kirgizib*, *esi kirsin*, *esing kirdi*, *esi kirib*, *esimni kiritib*, *esi kirar-chiqar*, *esi kirari-chiqari*, *esing kiradi*, *esini kirgizib*, *esi kirmasdan*, *esini kiritolmasalar-da*, *esi kirarli-chiqarli*, *esi kiradi*, *esimizni kiritmoqchi*.

### **О методе n-грамм**

Метод n-грамм может служить оптимальным инструментом для поиска и анализа устойчивых языковых единиц. Однако для достижения точного результата и повышения точности отображения результатов поиска таких единиц в корпусе важно использовать и другие методы и подходы в комбинации с ним. Использование лингвистического корпуса для создания базы данных фразеологических единиц является мощным и современным подходом, который не только позволяет выявить яркие образцы фразем в языке, но и создаёт полезный ресурс для преподавания языка и проведения исследований.

Метод n-грамм является весьма эффективным инструментом для обнаружения устойчивых языковых единиц, таких как фразеологические единицы, идиомы, афоризмы, а также других часто употребляемых выражений.

Использование n-грамм в корпусе для выявления фразеологических единиц представляет собой мощный и действенный метод, позволяющий определять устойчивые выражения на основе частотного анализа. Однако важно помнить, что модели n-грамм часто требуют дополнительной обработки и анализа для повышения точности выделения фразеологических единиц.

## **ЗАКЛЮЧЕНИЕ**

1. В корпусной лингвистике выделяются три типа работы с корпусами: (1) корпусо-ориентированные исследования; (2) исследования, основанные на корпусе; (3) исследования, использующие корпус как иллюстративный источник. В корпусной лингвистике фразеологизмы, отражённые в корпусах, могут служить либо лингвистическим материалом, либо методом анализа. При этом фразеологические единицы выступают как объект анализа, а языковые корпуса - как инструмент исследования. Активно развиваются исследования по тегированию фразеологизмов в корпусе, их представлению, разработке поисковых систем, а также методам извлечения и анализа фразем с помощью корпуса.

2. Фразема - это устойчивое выражение переносного значения, состоящее из двух и более лексем, которое по форме может быть эквивалентно словосочетанию или предложению. Для качественного тегирования фразем в корпусе необходимо указать их лексико-семантические, морфологические, грамматические, стилистические и другие характеристики в лингвистической базе данных.

3. Грамматическое варьирование может происходить за счёт замены, опущения или добавления грамматических компонентов. Устойчивое выражение может иметь множество парадигматических вариантов. В корпусе одному фразеологизму может соответствовать сотни грамматических или морфологических форм. Для фиксации всех вариантов необходимо наблюдать за их реализацией в большом корпусе текстов. Тегирование фразем в корпусах, рассчитанных на работу с большими данными, позволяет зафиксировать эти варианты.

4. В ряде фразем может изменяться морфологическая форма обеих компонентов. Однако не каждая морфологическая форма создает новый вариант выражения. Некоторые фраземы, отсутствующие в словаре, были обнаружены в корпусе в различных морфологических формах. Это делает возможным пополнение фразеологического словаря на основе данных корпуса.

5. Фразеологический слой узбекского языка богат и разнороден: некоторые фраземы активно употребляются, другие - крайне редко. Для количественной оценки частотности их употребления необходимо предварительное POS-тегирование фразем. В рамках проекта была создана база данных фразем, приведены их лемматизированные формы, и каждому фразеологизму присвоен тег (например, как существительное-фразема, глагольная фразема и т.д.).

6. Лингвистическая база данных - это структурированный набор сведений, содержащих слова, фраземы, грамматические формы, идиомы и иные единицы языка. Она строится на таблицах с определёнными полями, в которых фиксируются дефиниции, части речи, иллюстративные примеры и метаинформация. Этапы разработки включают: постановку цели, выбор платформы, проектирование структуры, ввод данных, разработку интерфейса, тестирование, документооборот, обновление и расширение.

7. Формирование базы данных включает: установление морфологической парадигмы фраземы, определение её вариантов, проверку в словаре, добавление при отсутствии, указание синонимов и завершение записи. Все фраземы классифицируются по частотности: низкой, средней и высокой.

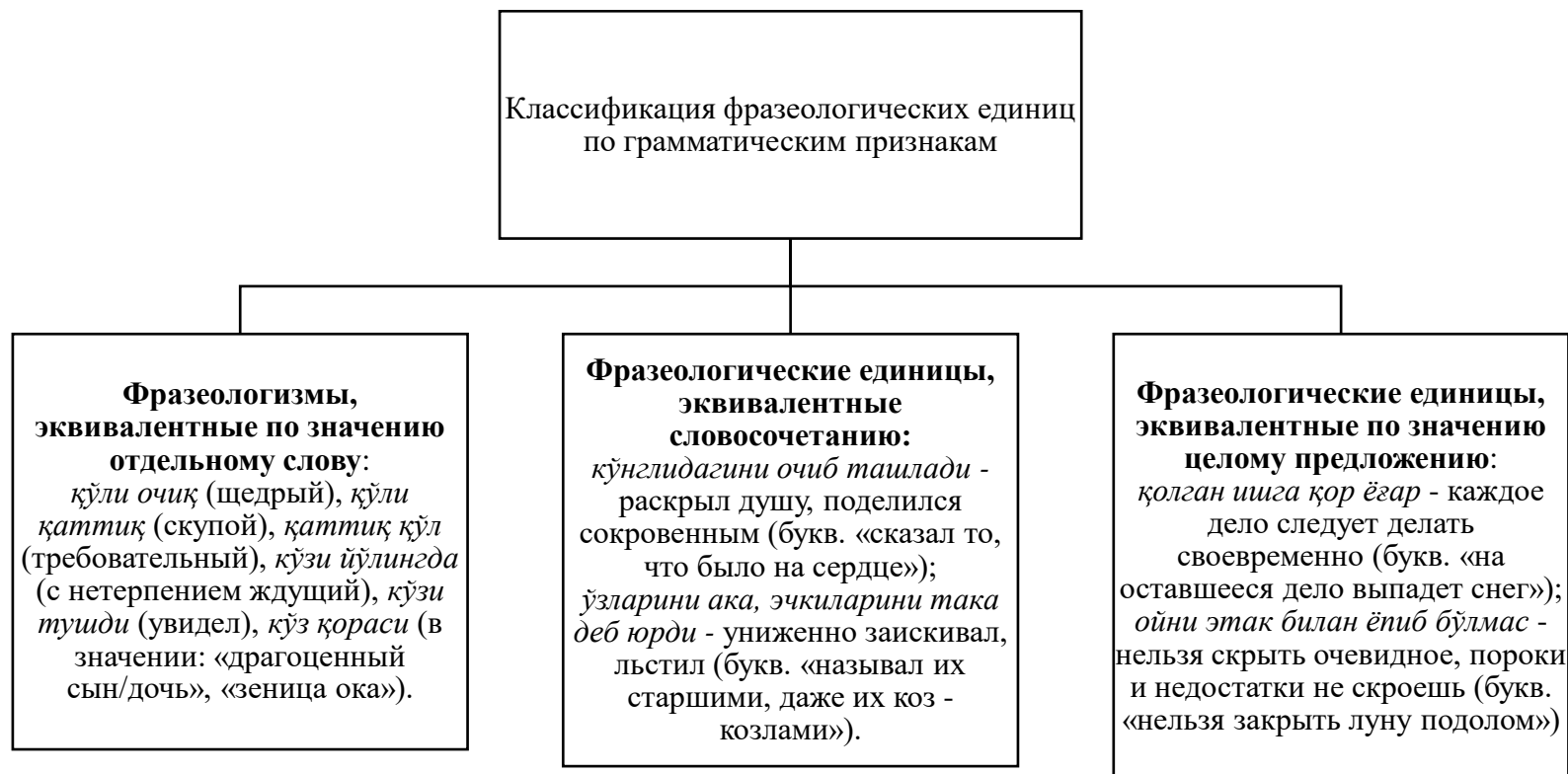
8. В узбекском языке существует омонимия как между различными фраземами, так и между фраземами и свободными словосочетаниями. При автоматическом извлечении они попадают в один список, что требует ручной верификации. Для автоматического решения этой проблемы предлагается использование n-граммных моделей.

9. Метод n-грамм эффективен при поиске устойчивых языковых единиц (фразем, идиом, афоризмов и часто употребляемых выражений). Этапы включают: генерацию n-грамм, частотный анализ, определение длины, анализ устойчивости и контекста, фильтрацию, кластеризацию, ручную проверку и использование на примерах.

10. Поиск фразем из словаря в корпусе даёт высокую точность и позволяет определить частотность. Благодаря морфологической нормализации и охвату грамматических вариантов (включая синонимы, времена глагола, личные формы), достигается высокая полнота поиска. Рекомендуются словарно-ориентированный подход в сочетании с дополнительными статистическими фильтрами. Перспективным направлением остаётся использование методов n-грамм, PMI, t-теста и машинного обучения для выявления новых устойчивых выражений, отсутствующих в словарях.

11. В результате исследования была составлена частотная таблица фразем, определены их стилистические свойства. Эти данные легли в основу тегирования корпуса с помощью специальных тегов (например, [IV]), что облегчает дальнейший поиск, фильтрацию и перевод.

12. Возможна семантическая и синтаксическая классификация фразем, изучение их распределения по жанрам. В перспективе можно исследовать использование фразем по сферам: например, провести сравнительный анализ фразем из политической речи и художественной литературы.



**Схема 1. Классификация фразеологических единиц по грамматическим признакам**

Asabni buzmoq		Avjga chiqmoq	
asabini buzgani-yu	asabim buziladi	avjga chiqqani	avjga chiqayapti
asabini buzib	asabni buzadi	avjga chiqadi	avjga chiqishi
asabim buzilib	asabim	avjga chiqqan	avjga chiqsa
asabini buzishdan	buzilganidanmi-	avjga chiqmoqda	avjga chiqdi
asablarini	yo‘qmi	avjga chiqdi	avjga chiqqanini
buzishlari	asablari buzilgani	avjga chiqishiga	<b>mavjudotga chiqarib</b>
asabni buzadigan	asabingni	avjga chiqqandi	<b>mavjudligi ochiqlandi</b>
asabi buziladi	buzmaslik	avjga chiqqanida	avjga chiqarishdi
asabi buzilgani	asablarim buzildi	avjga chiqib	avjga chiqqanligini
asabini buzadigan	asabim buzildi	avjga chiqaramiz	avjga chiqaveradi
asabimni buzdi	asabi buzilib	avjga chiqadigan	avjga chiqa
asabingizning	asab buzilishidan	avjga chiqarayotgan	avjga chiqaveradi
buzilgani	asabimizni	avjga chiqqanini	zavjalariga chiqarardi
<b>asabiy</b>	buzaverib	avjga chiqqan	Zavjangizga achchiq
<b>buzilishlar</b>	asabi buzildi	avjga chiqyapti	avji chiqib
asab buzilishi	asabi buzilishidan	avjga chiqqanda	avjga chiqqani
asabini buzadi	asabingiz buzilib	<b>mavjudotlar kirib-chiqib</b>	avjga chiqdiki
asab	asabini buzishni	avjga chiqishi	avjga chiqarkan
buzilishining	asabi buzilganligini	avjga chiqqan	avjga chiqardi
asab buzishiga	asab buzilishlari	avjga chiqayotganligidan	<b>Mavjuda chiqib</b>
asab buzilishiga	asabini buzari	avjga chiqarilgan	avjga chiqqanmish
asabingizni	asab buzilishlar	avjga chiqqanda	<b>mavjud chiqindilarni</b>
buzadi	asabini buzdi	avjga chiqishida	avjga chiqqach
asabini buzayapti	asabimni	avjga chiqqaniga	avjga chiqmoqda
asabni	buzganlaring	avjga chiqayotgan	avjga chiqmasa
buzmasdan	asablarimning	avjga chiqargan	avjga chiqqanining
asabi buzilgandan	buzilishi	avjga chiqyapti	<b>mavjlanib chiqqan</b>
asab buzilish	asabing buziladi	avjga chiqadi	<b>mavjudligi chiqindilar</b>
asabingiz	asabingizni	avjga chiqqanidan	avjga chiqarib
buzilishiga	buzishning	avjga chiqar	<b>mavjudligini ochiq</b>
asablari buzilib	asablarni buzishga	avjga chiqa	<b>mavjudot chiqib</b>
asabi buzilgan	asabingni	avjga chiqqanligi	<b>mavjudligi</b>
asabimizni buzib	buzmasdan	avjga chiqib	<b>ochiqchasiga</b>
	asabingizni	mavjud chiqindi	<b>mavjud ochiq</b>
	buzmang	avjga chiqar	
	asabini buzayotgan	avjga chiqmagandek	
	asabini buzishga	avjga chiqarayotgan	
	asablar	avjga chiqmoqchi	
	buzilishining	avjga chiqardi	
	asablar buzilishi		
	asablari buzilgan		
	asabini buzish		

Таблица 2. Низкочастотные и высокочастотные фразеологические единицы

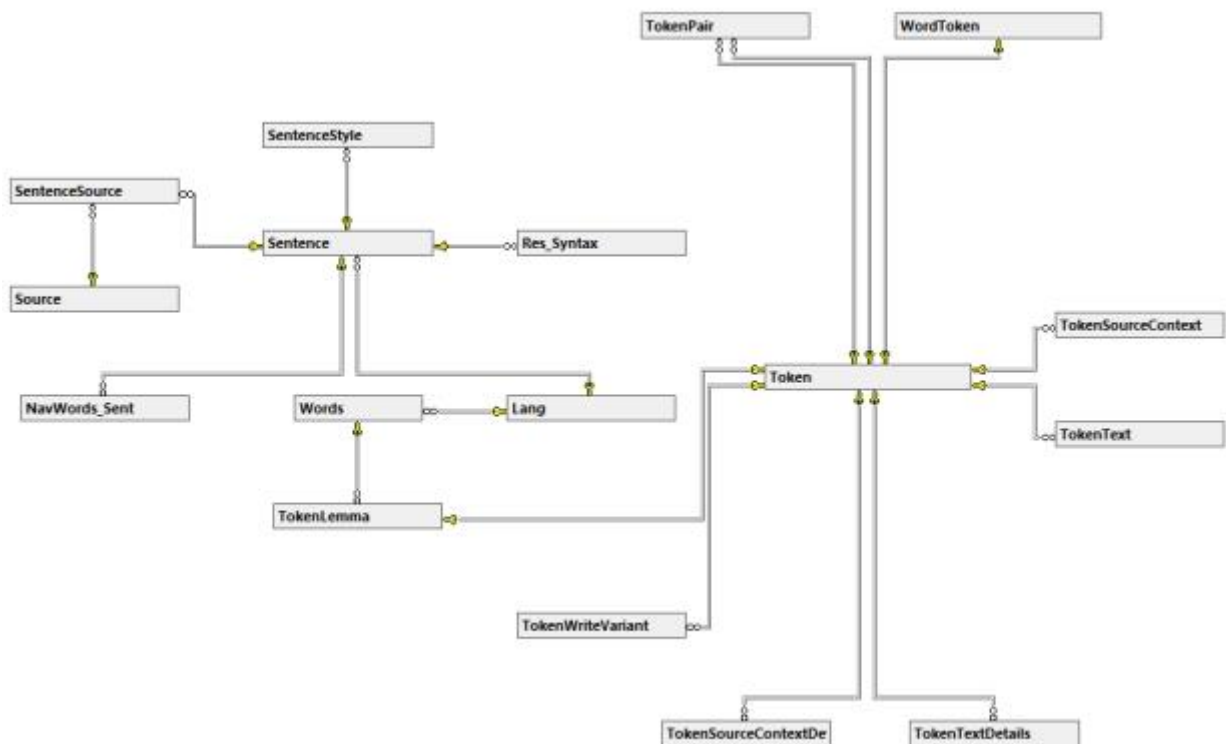


Схема 2. ER-модель для идентификации фразеологических единиц

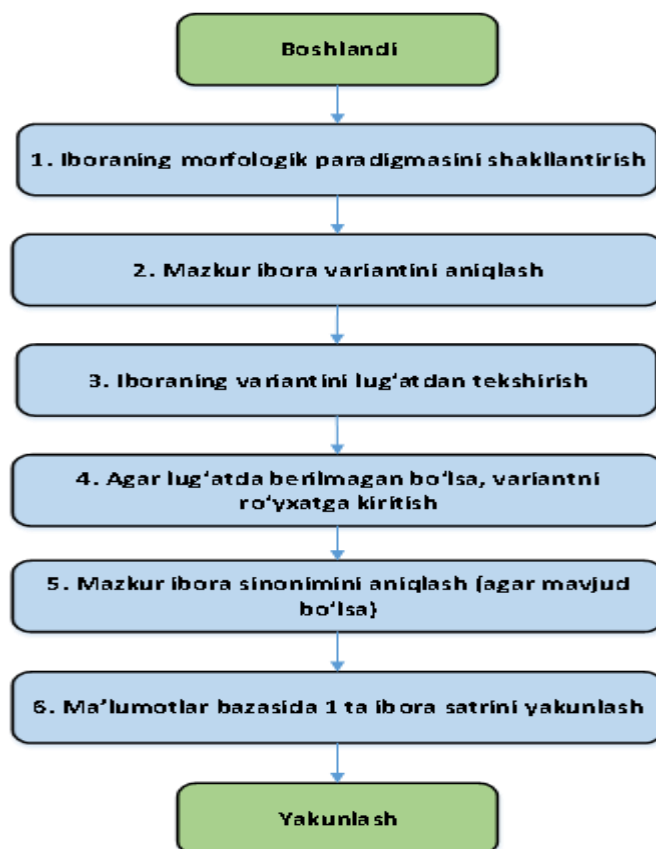
Bigram bo'yicha izlash

Natija: 24 ta yozuvdan 24 tasi ko'rsatilmogda.

**Bigram**

Ajali yetgan
ajaling yetgani
ajali yetgan
ajali yetib
ajali yetgudek
ajalim yetib
ajali yetmagan
ajali yetganda

Схема 3. Результаты поиска фразеологических единиц



**Схема 4. Процесс формирования базы данных**

**ONE-TIME SCIENTIFIC COUNCIL No. DSc.03/30.12.2019.Fil.19.01 FOR  
THE AWARD OF SCIENTIFIC DEGREES AT THE TASHKENT STATE  
UNIVERSITY OF UZBEK LANGUAGE AND LITERATURE NAMED  
AFTER ALISHER NAVOI**

**TASHKENT STATE UNIVERSITY OF UZBEK LANGUAGE AND  
LITERATURE NAMED AFTER ALISHER NAVOI**

**YODGOROV UMIDJON SAYDILLA UGLI**

**LINGUISTIC AND COMPUTATIONAL FOUNDATIONS OF TAGGING  
PHRASEOLOGICAL UNITS IN THE UZBEK LANGUAGE CORPUS**

**10.00.11 – Theory of Language. Applied and Computational Linguistics**

**ABSTRACT**

**Doctor of Philosophy (PhD) dissertations in philological sciences**

**Tashkent – 2025**

**The topic of the Doctor of Philosophy (PhD) dissertation is registered with the Higher Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan for B2023.3.PhD/Fil3963.**

The dissertation was completed at Tashkent State University of Uzbek Language and Literature named after Alisher Navoi.

The abstract of the dissertation in three languages (Uzbek, Russian, English (summary)) is posted on the website of the Scientific Council ([www.navoiy-uni.uz](http://www.navoiy-uni.uz)) and in the Information and Educational Portal “Ziyonet” ([www.ziyonet.uz](http://www.ziyonet.uz))

**Scientific supervisor:** **Elov Botir Boltayevich**  
Doctor of Philosophy (PhD) in Technical Sciences,  
Associate Professor

**Official opponents:** **Abdurakhmonova Nilufar Zaynabiddin qizi**  
Doctor of Philology, Professor

**Eshmuminov Askar Allamurodovich**  
Doctor of Philology, Associate Professor

**Lead organization:** **Samarkand State University**

The defense will take place on « \_\_\_\_ » \_\_\_\_\_ 2025 at \_\_\_\_\_ at the meeting of the one time Scientific Council DSc.03/30.12.2019.Fil.19.01 at the Tashkent State University of Uzbek Language and Literature named after Alisher Navoi (Address: 100100, Tashkent city, Yakkasaray district, Yusuf Khos Hajib street, 103. Tel.: (99871) 281–42–44; fax: (99871) 281–42–44, e-mail: [monitoring@navoiy-uni.uz](mailto:monitoring@navoiy-uni.uz)).

The dissertation can be reviewed at the Information Resource Center at Tashkent State University of Uzbek Language and Literature named after Alisher Navoi (registered under the number \_\_\_\_). Address: 100100, Tashkent city, Yakkasaray district, Yusuf Khos Hajib street, 103. Tel.: (99871) 281–42–44; fax: (99871) 281–42–44.

The abstract of the dissertation was distributed on « \_\_\_\_ » \_\_\_\_\_ 2025 y.  
(Protocol at the registred No. \_\_\_\_ on « \_\_\_\_ » \_\_\_\_\_ 2025.)

**Kh. Dadaboev**  
Chairman of the One-time Dissertation Council, established under the Scientific Council for the Award of Academic Degrees; Doctor of Philological Sciences, Professor

**K.U. Pardayev**  
Academic Secretary of the One-time Dissertation Council, established under the Scientific Council for the Award of Academic Degrees; Doctor of Philological Sciences, Professor

**S.B. Normamatov**  
Chairman of the Academic Seminar under the One-time Dissertation Council, established under the Scientific Council for the Award of Academic Degrees; Doctor of Philological Sciences, Professor

## INTRODUCTION (PhD) dissertation abstract)

**The aim of the research** is to develop the linguistic and computational foundations for tagging phraseological units of the Uzbek language within the corpus.

**The object of the research** is the phraseological units presented in the phraseological dictionaries of the Uzbek language.

**The scientific novelty of the research** is as follows:

as a result of studies conducted based on language corpora, experiments on the tagging of phraseological units were analyzed, and specific linguistic features of Uzbek phraseological expressions were identified. The structural components of phraseological units, their correlation with parts of speech, their fixed or variable forms, contextual behavior, and possibilities for automated recognition were thoroughly examined;

the practical significance of tagging phraseological units within a language corpus has been substantiated; linguistic approaches to the morphological annotation of Uzbek phraseological units have been defined; a special algorithm aimed at lemmatizing Uzbek phraseological expressions has been developed. Due to the creation of a phraseological database, opportunities for the automatic processing of phraseological units and their integration into linguistic resources have been expanded;

statistical and rule-based methods for the lemmatization of phraseological expressions have been developed, along with a specialized system of morphological tags for identifying and structurally analyzing phraseological units. The effectiveness of the statistical approach based on n-gram models for identifying phraseological units has been substantiated in practice;

furthermore, algorithms and rules for identifying grammatical variants of phraseological units using the n-gram method, as well as for detecting and eliminating homonymy in phraseological expressions, have been developed and applied to the Uzbek language corpus.

**Implementation of research results.** Based on the results obtained in the study during the tagging of the Uzbek phraseological system, the following implementations have been carried out:

as a result of research conducted using language corpora, tagging experiments on phraseological units were analyzed, and linguistic features specific to Uzbek phraseological units were identified, including their structural components, associations with parts of speech, fixed or variable forms, contextual realizations, and possibilities for automatic recognition. These scientific findings were implemented into the activities of the State Institution “Center for Teaching and Improving Skills in Official Language Usage” under the Alisher Navoiy Tashkent State University of Uzbek Language and Literature (based on the institution’s official record dated February 4, 2025). As a result, a database consisting of 1800 tagged Uzbek phraseological units was formed using the existing electronic resources within the framework of the Center for the Introduction of E-learning in educational institutions.

the practical relevance of tagging phraseological units in a language corpus was substantiated; linguistic approaches to the morphological tagging of Uzbek phraseological units, an algorithm specifically designed for their lemmatization, and a database of phraseological units enabling their automatic processing and integration into linguistic resources - these scientific conclusions were utilized in the project titled “UzUDT: Universal Dependency Treebank for the Uzbek Language and its Semantic Analysis”, implemented in 2022–2024 at the National University of Uzbekistan named after Mirzo Ulugbek under grant REP-25112021/113 (according to Reference №04/11-5273 dated April 23, 2025, provided by the university). As a result, based on the developed mathematical models, algorithms, and software, the possibility of accurately identifying phraseological units within sentence structures was achieved; using semantic concordance methods and algorithms, the syntactic analysis accuracy of the 1800 tagged phraseological units was increased to 92%.

scientific conclusions related to statistical and rule-based methods for lemmatizing phraseological units, a dedicated system of morphological tags for identifying and structurally analyzing them, and a statistical approach based on n-gram models for phraseological unit recognition were employed in the applied project “Creating an Educational Corpus of the Uzbek Language”, implemented in 2021–2023 at Alisher Navoiy Tashkent State University of Uzbek Language and Literature under grant IZ-2021020813 (according to Reference №04/1-4250 dated December 24, 2024). As a result, the lemmatization algorithm and the phraseological unit database were applied in the development of the search system for the Uzbek language educational corpus, and the algorithm was successfully used to detect and resolve homonymy within phraseological units.

an algorithm for identifying grammatical variants of phraseological units based on the n-gram method, an algorithm for detecting homonymy in phraseological units, and the corresponding rule sets were utilized within the same applied project “Creating an Educational Corpus of the Uzbek Language”, carried out in 2021–2023 at Alisher Navoiy Tashkent State University of Uzbek Language and Literature under grant IZ-2021020813 (according to Reference №04/1-4250 dated December 24, 2024). As a result, the n-gram-based search capability within the search system of the Uzbek language educational corpus was enhanced.

**The structure and scope of the dissertation.** The dissertation consists of an introduction, three chapters, a conclusion and a list of used literature, the total volume of the dissertation is 124 pages.

**E'LON QILINGAN ISHLAR RO'YXATI**  
**СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**  
**LIST OF PUBLISHED WORKS**

**I bo'lim (I часть; I part)**

1. Yodgorov U.S. O'zbek tilshunosligida frazeologizmlarning o'rganilishi, lisoniy xususiyatlari va nutqiy qiymati // O'zbekiston Milliy universiteti xabarlari, 2023, 1/11, 277-280-b. (10.00.00, №15)
2. Yodgorov U.S. Til korpusida iboralarni teglashning lingvistik asoslari va ma'lumotlar bazasi // O'zbekiston Milliy universiteti xabarlari, 2023, 1/11, 348-350-b. (10.00.00, №15)
3. Yodgorov U.S. O'zbek tili korpusi asosida frazeologik birliklar ma'lumotlar bazasini shakllantirish. // O'zbekiston Milliy universiteti xabarlari, 2024, 1/11, 286-289-b. (10.00.00, №15)
4. Yodgorov U.S. O'zbek tili frazemalarining morfologik shakli va variantlari tadqiqiga an'anaviy va korpus tahlili yondashuvi // O'zbekiston: til va madaniyat 2024, 3, 29-48-b. (10.00.00, OAK Rayosatining 2021 yil 30 oktabrdagi 308/6-son qarori)
5. Yodgorov U.S. Formation of a database of phraseological units based on the corpus of the Uzbek language // American Journal of Philological Sciences ISSN – 2771-2273 Vol.05 Issue 03 2025. (Researchbib, IF=10.36)
6. Yodgorov U.S. The Significance of Tagging Phraseological Units in the Language Corpus // Current Research Journal of Philological Sciences ISSN: 2767-3758 Vol.06 Issue 03 2025. (Researchbib, IF=12.07)
7. Yodgorov U.S. Elektron lug'at axborot-qidiruv tezarusi va ontologiyalarni ishlab chiqish vositalari // "Kompyuter lingvistikasi: muammolar, yechim, istiqbollar" xalqaro ilmiy-amaliy konferensiya, 2022, 431-442-b.
8. Yodgorov U.S. Text generator based on Markov chains // Proceedings of International scientific conference Statistics and its applications. 19-20 October, 2022, 438-441-b.
9. Yodgorov U.S. Text generator based on lexis-syntactic templates // "Kompyuter lingvistikasi: muammolar, yechim, istiqbollar" respublika ilmiy-amaliy konferensiya, 2023, 173-179-b.
10. Ёдгоров У., Эркинов Ф., Юлдашов А. Способ сгенерировать текст с помощью цепи Маркова. // VI Международная научно-практическая конференция Статистика и её применения октябрь 19-20, 2022, 435-437-b.
11. Yodgorov U.S. O'zbek tili korpusi asosida frazeologik birliklar ma'lumotlar bazasini shakllantirish bosqichlari va ER sxemasi. // "Kompyuter

lingvistikasi: muammolar, yechim, istiqbollar” Respublika I ilmiy-texnikaviy konferensiya, 2021, 270-276-b.

12. Yodgorov U.S. O‘zbek tili frazeologik birliklarida omonimiyani ngramm usulida aniqlash. // “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” Respublika I ilmiy-texnikaviy konferensiya, 2021, 263-269-b.

## **II бўлим (II часть; II Part)**

13. Elov B., Alayev R., Xusainova Z., Yodgorov U. CBOW neyron tarmoqlari vositasida o‘zbek tili so‘zlarini bashoratlash // Xorazm Ma’mun akademiyasi axborotnomasi, 2024, №2/4, 48-57-b. (10.00.00, №21)

14. Elov B., Yodgorov U., Yuldashev A. Lingvistik korpuslar tipologiyasi va ularga qo‘yilgan talablar // Xorazm Ma’mun akademiyasi axborotnomasi 2024, 12/4, 125-130-b. (10.00.00, №21)

15. Elov B., Yodgorov U., Yuldashev A. Til korpusi turlari va umumiy xususiyatlari // Xorazm Ma’mun akademiyasi axborotnomasi 2024, 12/4, 119-124-b. (10.00.00, №21)

16. Elov B., Khamroeva Sh., Alayev R., Xusainova Z., Yodgorov U. Methods of processing the uzbek language corpus texts // International Journal of Open Information Technologies ISSN: 2307-8162 vol. 11, no. 12, 2023. (Researchbib)

17. Yodgorov U. Ijtimoiy-iqtisodiy tadqiqot natijalari ishonchliligini statistik usullar yordamida baholash // O‘zbekiston Milliy universiteti xabarлари, 2022, 1/11/1, 77-79-b. (10.00.00, №15)

18. Elov B., Xusainova Z., Yodgorov U.S. O‘zbek tili matnlari uchun tokenizayorni ishlab chiqish // Muhammad Al-xorazmiy avlodlari, 2023, 2(24), 27-33-b. (05.00.00, №10)

19. Elov B., Hamroyeva Z., Alayev R., Xusainova Z., Yodgorov U. O‘zbek tili korpusi matnlarini qayta ishlash usullari // Raqamli Transformatsiya va Sun’iy Intellekt, Vol.1, Issue 3, 2023, 117-129-b. (05.00.00, OAK Rayosatining 2023 yil 4 iyuldagi 340/5-son qarori)

20. Elov B., Hamroyeva Sh., Xusainova Z., Yodgorov U., Xudayberganov N., A.Yuldashev Pos tagging of uzbek texts using hidden markov models (HMM) and viterbi algorithm // Materials of International Scientific and Practical Conference on "The Role of the National University of Uzbekistan in the Development of Science and Social Progress" (Part I), May 12, 2023, 104-115-b.

21. Yodgorov U.S. Assessment of reliability of the results of socio-economic research by statistical methods // Materials of International Scientific and Practical Conference on "The Role of the National University of Uzbekistan in the Development of Science and Social Progress" (Part I), May 12, 2023, 424-427-b.

22. Elov B., Hamroyeva Sh., Xudayberganov N., Yodgorov U.S, Yuldashev A. POS tagging of Uzbek texts using Hidden Markov Model (HMM) and Viterbi algorithm // IMS International Workshop «Interactive Systems & Information Society Technologies – InterSys2023». Санкт-Петербург, 26-28 июня 2023 г.

23. Elov B., Hamroyeva Sh., Yodgorov U., Matyakubova N. One-hot encoding and bag-of-words methods in processing the uzbek language corpus texts // Seventh International Conference on Computer Processing of Turkic Languages. Turklang 2023, 639-645-b.

24. Elov B., Yodgorov U. Elektron lug‘atlar, tezarus, ontologiyalar elektron ta’limda zaruriy vositalar sifatida // “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” xalqaro ilmiy-amaliy konferensiya, 2021, 177-182-b.

25. Elov B., Alayev R., Yodgorov U. O‘zbek tili frazeologik birliklarini dasturiy ta’minoti. O‘zbekiston Respublikasi Adliya vazirligi, DGU 31229, 07.12.2023.

26. Elov B., Alayev R., Yodgorov U. O‘zbek tili frazeologik birliklarining lingvistik bazasi. O‘zbekiston Respublikasi Adliya vazirligi, DGU 31230, 07.12.2023.