

ЎЗБЕКИСТОН РЕСПУБЛИКАСИ ФАНЛАР АКАДЕМИЯСИ
АЛИШЕР НАВОИЙ НОМИДАГИ ТИЛ ВА АДАБИЁТ ИНСТИТУТИ

Кўлёзма ҳуқуқида
УДК 811.512.133'33

РИЗАЕВ СУННАТУЛЛА АБДУМАННАПОВИЧ

**ЎЗБЕК ТИЛИНИНГ ЛИНГВОСТАТИСТИК
ТАДҚИҚИ**

10.02.01 - ўзбек тили

Филология фанлари доктори илмий даражасини олиш
учун тақдим этилган диссертация

А В Т О Р Е Ф Е Р А Т И

ТОШКЕНТ – 2008

1194.3 (043.3)

Иш Тошкент давлат иктисодиёт университети Ўзбек ва рус тиллари ҳамда Ўзбекистон Миллий университети Умумий тилшунослик ва компьютер лингвистикаси кафедраларида бажарилган

Илмий маслаҳатчилар: филология фанлари доктори, профессор-математик
БЕКТАЕВ Қалдибай Бектаевич

филология фанлари доктори, профессор
ДАДАБОЕВ Ҳамидулла Арипович

Расмий оponentлар: филология фанлари доктори, профессор
КАРИМОВ Суюн Каримович

филология фанлари доктори, профессор
НУРМОҢОВ Абдулхамид Нурмонович

филология фанлари доктори, профессор
АЙЫМБЕТОВ Максат Каллыевич

Этакчи ташкилот: Тошкент давлат педагогика университети

Ҳимоя Ўзбекистон Республикаси Фанлар академияси Алишер Навоий номидаги Тил ва адабиёт институти ҳузуридаги ДК 015.04.02 рақамли кенгашнинг 2008 йил "19" декабрь соат 14 да ўтадиган мажлисида бўлади.
Манзил: 700170, Тошкент шаҳри, И.Мўминов кўчаси, 9-уй.

Диссертация билан Ўзбекистон Республикаси Фанлар академиясининг Асосий кутубхонасида танишиш мумкин.

Автореферат 2008 йил "10" ноябрь да тарқатилди.

Ихтисослашган кенгаш
илмий котиби


Н. М. МАҲКАМОВ

1. ДИССЕРТАЦИЯНИНГ УМУМИЙ ТАВСИФИ

Мавзуининг долзарблиги. Тараққиётнинг ҳар бир босқичи – бу янги муаммолар ва уларни бартараф этиш йўлидаги янги вазифалар демокдир¹.

XXI аср – компьютер² технологиялари, илгор педагогика ва ахборот технологиялари (интернет), юксак тафаккур, фан ва техниканинг жадал тараққиёт асри бўлади. Аср бошига келиб “Электрон таълим”, “Электрон бошқарув”, “Электрон ҳукумат”, “Очиқ таълим”, “Масофавий таълим”, “Ахборотлашган таълим” каби тушунчалар ҳаётимизга сингиб бормоқда. Ўзбекистон Республикаси Президенти Ислоҳ Каримов таъкидлаганларидек: “Ҳозирги ахборот, коммуникация ва компьютер технологиялари асрида, интернет кундан-кунга ҳаётимизнинг барча жабхаларига тобора чуқур ва кенг кириб бораётган бир пайтда, одамларнинг онги ва тафаккури учун кураш ҳал қилувчи аҳамият касб этаётган бир вазиятда бу масалаларнинг жамиятимиз учун нақадар долзарб ва устувор бўлиб бораётгани ҳақида гапиришга ўтиришга ҳоҷат йўқ”³.

Ҳозирда инжнерлик лингвистикаси, ҳисоблаш лингвистикаси, математик лингвистика ва “электрон - таржимон”, компьютер лингвистикаси юзасидан бир қатор янги муаммолар ўртага ташланмоқда. Зеро замонавий компьютер жуда улкан имкониятларга эгаки, у учун шахмат ўйнашдан тортиб, тиббий ташхис қўйиш, корхоналарни бошқариш, қадимги халқларнинг сирли хатларини “ўқиш”, мазмунини сўзлаб бериш, шеър ёзиш, қуйлар ижро этиш, катга ҳажмдаги китобларга аннотация ёзиш, лойиҳалар тарҳини чизиши каби юзлаб янгиликларнинг барчасида, албатта, замонавий компьютер технологиялари ҳам иш кучи, ҳам вақт жиҳатидан жуда яқин ёрдамчи ҳисобланади (одам + машина + одам; одам + матн + машина + одам).

Бундай илмий тадқиқотлар сирасига мураккаблиги, долзарблиги билан характерланадиган соҳалардан бири ҳисобланмиш лингвостатистика ҳам ҳосдир. Лингвостатистикага оид дастлабки ишлар гарчи миқдорий тасвирлаш характерида бўлса ҳам, булар кейинчалик назарий ҳулосаларга замин яратади. Ҳар қандай фанда бўлганидек, лингвостатистикада ҳам кейинги, бирмунча мураккаб қадам унинг назариясини яратиб, “лингвистик тенглама”ларни тузиш, у ёки бу қонуният миқдорининг сифат жиҳати билан узвий алоқада олдидан айтиб беришдан иборат бўлади.

Бир-бирларидан бирмунча узоқ бўлган турли фан соҳаларининг ўзаро алоқаралашуви (интеграцияси) боис, янги-янги тажрибалар, ҳулоса ва муаммолар

¹ Каримов И. Биз таялган йўл — демократик тараққиёт ва мўърифий дунё билан ҳамкорлик йўли.— Т.: Ўзбекистон, 2003. Т. XI. — Б.14.

² Инглизча computer сўзи лотинчадаги computo, computate сўзидан олинган бўлиб, “ҳисоблайман”, “санайман” деган маънони билдиради. Инглиз тили адабиётларида электрон ҳисоблаш машинаси ўрнида қўлланаб келинмоқда. Қараг: Hojiyev A., Nurmooov A., Zaynobiddinov S., Kokren K., Saidxonov M., Sobirov A., Quronov D. Hozirgi o'zbek tili faol so'zlarining izohli lug'ati. – T.: Sharq, 2001. – B. 111., Ўзбек тилининг изоҳли лугати (5 жылдамик). – Т.: ЎМЭ, 2006. 2-жылд. – Б.397.

³ Каримов И. Бизнинг бош мақсадимиз – жамиятни демократлаштириш ва янгилаш, мамлакатни модернизация ва ислоҳ этишдир. Ўзбекистон демократик тараққиётнинг янги босқичида. – Т.: Ўзбекистон, 2005. – Б.38-39.

келиб чиқмоқда. Хусусан, ўзбек тилшунослиги олдидаги янги-янги илмий-амалий муаммоларни оқилона ҳал этиш, лингвистик ҳодисаларни аниқ, объектив, қатъий ва ишончли тадқиқ этиш, тасвирлаш учун математик-статистик методлар, ахборот технологиялари билан яқиндан алоқани йўлга қўйиш, мустаҳкамлаш ва такомиллаштириш муҳим ҳисобланади.

Мазкур тадқиқотнинг таркибий қисмларидан бири бўлган ҳарфий статистика ўзбек тили ҳарфлари билан боғлиқ барча ахборот технологияларини жадаллик билан ривожлантириш имкониятини яратади. Фонологик системанинг лингвостатистик тадқиқоти асосида яратилган лингвостатистик параметр (қонуният)лар нафақат қиёсий-типологик ва информацион-статистик изланишларда шунингдек, тилнинг қай даражада оҳангдорлигини белгилашда ҳам қўл келади. Бўғиннинг статистик структураси эса ўзбек тили сўз таркибининг неча бўғиндан иборат эканлигини аниқлаш билан бирга дефектологиянинг логопедия соҳаси учун манба бўлиб хизмат қилади. Сўз морфем структурасининг лингвостатистик тадқиқ этилиши натижаси морфемалар сонининг сўзшакллари сонига муносабати тилнинг синтетиклик индексини кўрсатади. Лексик (сўзшакллар) статистикаси эса маъна фонди (лингвистик банк)нинг яратилишига имкон беради.

Ўзбек тилшунослигида тил ёки нутқнинг график, фонологик, бўғин, морфем ва лексик структурасини математик-статистик ва информацион-типологик методлар асосида тадқиқ ва таҳлил этиш ҳамда уларнинг натижасида эришилган лингвостатистик қонуният(параметр)лар назарий ва амалий тилшуносликка ўта муҳим материал бўла олиши танланган мавзунинг долзарблигини белгилайди.

Муаммонинг ўрганилганлик даражаси. Тилшуносликнинг сўнги тадқиқотларига қўра тил ва нутқ¹ бирликларининг тадқиқи ва таҳлилини замонавий ахборот технологияларисиз тасаввур этиб бўлмайди. XX асрнинг 60-йилларига келиб, лингвостатистикада Пьер Гиро²нинг “Лингвостатистика методи ва муаммолари” ҳамда Густав Хердан³нинг бир қатор асарлари муҳим аҳамиятга эга бўлди.

П.Гиро библиографиясининг “Эсту-Циф тенгламаси”, “Луғатнинг статистик характеристикаси”, “Рембо стилининг эволюцияси”, “Шеърнинг фонетик структураси” каби бўлимларидан филологик тадқиқотлар учун манбалардан бири сифатида фойдаланиш мумкин.

¹ Тил бирликлари бошқа-ю, нутқ бирликлари бошқа. Тил хатоси бошқа-ю, нутқ хатоси бошқадир(Ҳожиев А. Лексема лисоний бирликдан ёки нутқ бирлигига? // Ўзбек тили ва адабиёти. – Тошкент, 2004. №1 – Б.49-54; Тил қурилишининг асосий бирликлари юзасидан мулоҳазалар // Ўзбек тили ва адабиёти. – Тошкент, 2004. №3. – Б.29-36; Незьмитов А., Бозоров О. Тил ва нутқ. –Т.: Ўқитувчи, 1993).

² Guiraud P. Problems et methodes de la statistique Linguistique. - Paris, 1960; Guiraud P. Bibliographie critique de la statistique Linguistique. - Utrecht, 1954 (қарийб 2500га яқин номни ўз ичига олади). Д.М. Сегалнинг маълумот беришича, П.Гиро библиографиясининг қўлёзмаси Прага Давлат стенография институтида сақланади ва унинг баъзи махсус стенографик журналда жойлашган: Тенорисник Rozhledy. - Praha, 1924.

³ Herdan Gustav. Language as Choice and Chance. - Groningen, 1956; id. Type - token mathematics. - The Hague, 1960; id. The Calculus of Linguistic observations. - The Hague, 1962; id. Quantitative Linguistics. - L., 1964; id. The Advanced Theory of language as choice and chance. - Berlin, 1966.

Ўтган асрнинг 70-йилларида “Нутқ статистикаси” гуруҳининг раҳбари, профессор Р.Г. Пиотровский ва профессор-математик Қ.Б.Бектаев ҳамкорлигида Чимкент Кимё технология институтининг “Машина таржимаси” лабораториясида кимё-полимерлар соҳаси бўйича биринчилардан бўлиб саноатлаштирилган машина таржимаси амалга оширилган эди¹.

Н.Ёқубованинг 1974 йилдаги номзодлик диссертациясида ўзбек сўзшакллари синтетлашнинг формал тасвири ва икки компонентли сўз бирикмаларида синтактик аффикслар ўринлашувининг формал қоидалари келтирилган. С.У.Мусахўжаевнинг “Лингвостатистический анализ узбекского научно-технического текста и терминологии подязыка электротехники и электроники” (1986) номли номзодлик диссертациясида ўзбек илмий-техника матнлари лексикасини статистик моделлаштириш ва ўзбек илмий техника терминологиясини лингвостатистик тадқиқ этишга эътибор қаратилган ҳамда 2-4 бирикмали терминлар аниқланган. С.Мухамедовнинг “Статистический анализ лексико-морфологической структуры узбекских газетных текстов” номли номзодлик иши (1980)да газета матнларининг лексик-морфологик тавсифи берилган. М.К.Айымбетов “Проблемы и методы количественно-типологического измерения близости тюркских языков (1997) номли докторлик диссертациясида қорақалпоқ, қозоқ ва ўзбек тили материаллари юзасидан тадқиқот олиб борган. А.Иброҳимовнинг “Бобур “Девон”и тилидаги ўзлашма лексика” (2006) номли монографиясида “Девон”нинг лексикаси частотасига кўра статистик, семантик ва этимологик таҳлили ҳамда “Девон”га конкорданс тузиш принциплари баён этилган. С.Мухамедованинг 2006 йилда чоп этилган методик қўлланмасида эса ҳаракат феъллари асосида компьютер дастурлари учун лингвостатистик таъмин яратишга оид фикр-мулоҳазалар билдирилган². С.Каримов, А.Қаршиев, Г. Исроиловаларнинг “Абдулла Қаҳҳор асарлари тилининг алфавит-частотали, частотали ва терс лугати” (2007)да лингвостатистик тадқиқ масалаларига ҳам эътибор қаратилган. Бироқ ҳозирга қадар тил сатҳларининг узвийлиги асосида компьютер ёрдамида комплекс лингвостатистик тадқиқ этилмаган. Бундай муаммоларнинг замонавий техникалар ёрдамида лингвостатистик тадқиқ этилиши юқорида таъкидланганидек назарий ва амалий тилшуносликдаги ечимини кутаётган қатор масалалар диққат эътиборни янада яқинроқ жалб этади.

Диссертация ишининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация мавзуси 1975 йилда ҳозирги Жаҳон Тиллари университети (собиқ Ф.Энгельс номидаги Тошкент давлат чет тиллар педагогика институти) Умумий тилшунослик кафедрасида (профессор-математик Қ.Б. Бектаевларнинг илмий маслаҳатчилигида) маълумланган ва институт Илмий Кенгаши томонидан тасдиқланган. Қ. Б. Бектаев вафот этгач, ўтган давр

¹ Пиотровский Р.Г., Бектаев К.Б. Машинный перевод: теория, эксперимент, введение // Вопросы языкознания – М., 1977. - №5. – С.27-33.

² Мухамедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. -Т., 2006. Po'latov A., Mubamedova S. Kompyuter lingvistikasi. — Toshkent, 2008. – 104 б.

мобайнида мазкур тадқиқот Мирзо Улугбек номидаги ЎзМУ Умумий тилшунослик ва компьютер лингвистикаси кафедрасининг “Тилнинг ижтимоий, тарихий ва диалектик тараққиёти” мавзуси бўйича олиб борилаётган илмий тадқиқот ишларига мос келадиган ва у ф.ф.д., профессор Ҳ.А.Дадабоевнинг илмий маслаҳатчилигида қайта кўриб чиқилиб, тузилган ва янги ашёвий материаллар билан тўлдирилиб, 1999 йил 24 июнда Тошкент давлат иқтисодиёт университети Илмий кенгашида қайта тасдиқланган.

Тадқиқот мақсади. Ўзбек тили матн материаллари асосида ҳарфлар, фонемалар, бўғинлар, морфемалар ва лексема(сўзшакл)лар қўлланиш частотасининг лингвостатистик тадқиқ ва таҳлил этиш пировардида лингвостатистик қонуният(параметр)ларни ишлаб чиқиш мазкур тадқиқотнинг асосий мақсади ҳисобланади.

Тадқиқот вазифалари. Диссертацияда қуйидаги вазифаларни бажариш режалаштирилди:

- статистик танланмаларнинг ташкил этиш хусусиятларини аниқлаш;
- ҳозирги ўзбек тили матнларида ҳарфлар қўлланиш частотаси фоизини белгилаш, статистик ишлаб чиқиш ва шулар асосида лингвостатистик параметрлар яратиш;
- ҳозирги ўзбек тилининг босма ҳарфлари энтропиясини илк бор аниқлаш;
- фонемаларнинг частотаси ва “фонематик занжир”даги фонемаларнинг микдорий муносабатларини белгилаш;
- ҳозирги ўзбек тилининг турли функционал стилларидаги фонемаларнинг қўлланиш частотаси ва “фонематик занжир”даги фонемаларнинг ўзаро микдорий муносабатларини аниқлаш ва лингвостатистик параметрлар яратиш;
- бўғин ва бўғин турларининг ҳамда бўғин типларининг статистик структурасини аниқлаш ва лингвостатистик параметрлар яратиш; бўғин таркибидаги унли ва ундош фонемаларни лингвостатистик таҳлил этиш;
- сўз(сўзшакл)ларнинг морфем(атик) структурасини аниқлаш;
- болалар адабиёти тили лексикаси ва уларнинг ўзига хос хусусиятларини лингвостатистик тадқиқ ва таҳлил этиш;
- Ҳ. Ҳ. Ниязий “Девон”и тили лексикаси (сўз шакллари)нинг статистик структураси ва лингвостатистик параметрларни аниқлаш, лингвостатистик тадқиқ ва таҳлил этиш, конкорданс тузиш.

Тадқиқот объекти ва предмети. Ўзбек тили турли функционал стиллари оригинал материаллари (проза, поэзия, публицистика, фалсафа, тарих, физика ва математика), болалар адабиёти тили лексикаси, шунингдек, Абдулла Қаҳҳорнинг “Синчалак” повести ҳамда Ҳамза Ҳакимзода Ниязий “Девони” тили лексикаси ва конкорданси асосида ҳарфий статистика, фонологик статистика, бўғин статистикаси, морфем статистика ва лексик статистика талқинида тадқиқ ва таҳлил этилди.

Тадқиқот методлари. Тадқиқотнинг методологик асосини тил ва нутқ ходисалари билан боғлиқ бўлган диалектик таълимот(сон ва сифат, шакл ва

мазмуни), статистик ва қиёсий-типологик метод ташкил этади. Бунда, айниқса, диалектиканинг миқдор ўзгаришлари ҳам сифат ўзгаришларига ўтиш ҳолатларининг рўй бериши, статистик метод асосида янги-янги белгиларнинг, кирраларнинг очила бориши, тадқиқотда таҳлил ва синтезнинг мутаносиблиги (уйғунлиги) ҳамда билиш назарияси, фан ва амалиётнинг ўзаро алоқадорлигининг органик тамойили, математик лингвистика ва компьютер лингвистикаси, лингвостатистик, тавсифий, қиёсий-типологик, систем-структур, моделлаштириш, статистик-информацион таҳлил методлари асос ҳисобланади. Шунингдек, Р. Г. Пиотровский бошчилигидаги юздан зиёд тилшунос, муҳандис ва математиклар иштирокидан 1959 йилда ташкил топган “Ўтқ статистикаси” гуруҳининг ленинградлик, белорусиялик, доғистонлик, ўрта осиелик ва молдавиялик жамоаларнинг барча автоматлаштирилган лингвостатистик тадқиқотлари йўналиши бўйича бажарилган ишлари мазкур диссертация учун илмий-назарий ва амалий манба вазифасини ўтади.

Ҳимояга олиб чиқилаётган асосий ҳолатлар:

- тил ва ЭҲМ (компьютер) уйғунлигининг оламшумул имкониятлари кўрсатиб берилди;

- ўзбек тили сатҳлари (харф, фонема, бўғин, морфема ва лексема) замонавий методлар асосида узвийликда лингвостатистик тадқиқ ва таҳлил этилиб, мазкур соҳада бажарилиши зарур бўлган айрим масалалар ўртага ташланди;

- тадқиқотга доир хорижий ва ўзбек тилшунослигига хос лингвостатистик тадқиқотлар таҳлил этилди;

- лингвостатистик танланмаларнинг хусусиятлари очилди;

- илк бор ўзбек матнларида ҳарфлар қўлланиш частотаси ва энтропияси ишлаб чиқилди;

- ўзбек тили ҳарфлари ва фонемаларининг лингвостатистик параметрлари, ўзбек тилининг сонанглилик даражаси, шунингдек, икки фонемали бирикмалар, бўғинлар, морфемалар, сўз(сўзшакл)ларнинг лингвостатистик параметрлари ишлаб чиқилди ва истиқболлари белгиланди. Шунингдек, лингвостатистик тадқиқ натижалари шу йўналишдаги бошқа тадқиқотлар билан қиёсий-қоғиштирма ва статистик-типологик аспектда ўрганиш имкониятларига йўл очилди ҳамда булар ўзбек тили сатҳлари билан боғлиқ ахборот технологиялари ҳамда компьютер лингвистикаси учун муҳим илмий-назарий ва амалий манба вазифасини ўтади.

Ишнинг илмий янгилиги:

- статистик танланмаларни ташкил этиш хусусиятлари ишлаб чиқилди;

- ҳозирги ўзбек тили матнларида ҳарфлар қўлланиш частотаси аниқланди, статистик ишлаб чиқилди ва шулар асосида лингвостатистик параметр (қонуният)лар белгиланди;

- ҳозирги ўзбек тилининг босма ҳарфлари энтропияси илк марта аниқланди;

- ҳозирги ўзбек тилининг турли функционал стилларидаги фонемаларнинг қўлланиш частотаси ва “фонематик занжир”даги фонемаларнинг ўзаро миқдорий муносабатлари аниқланди ва лингвостатистик параметрлар яратилди;

- ҳозирги ўзбек тилида жуфт қўлланадиган фонема (икки фонемали бирикма)лар қўлланиш частотаси аниқланди ва уларнинг лингвостатистик параметрлари яратилди;

- бугин ва бугин тилларининг статистик структураси белгиланди ва лингвостатистик параметрлари ишлаб чиқилди; бугин таркибидаги унли ва ундош фонемалар лингвостатистик таҳлил этилди;

- сўз(сўзшакл)ларнинг морфем(агик) структураси ишлаб чиқилди. Бунда морфемалар сонининг сўзшакллари сонига муносабати тил синтезининг ўлчови(синтетик индекси) бўлиб хизмат этиши асослаб берилди;

- болалар адабиёти тили лексикаси ва уларнинг ўзига хос хусусиятлари лингвостатистик тадқиқ ва таҳлил этилди;

- Ҳамза Ҳақимзода Ниёзий “Девон”и тили лексикаси (сўз шакллари)нинг статистик структураси ва лингвостатистик параметрлари ишлаб чиқилди.

Тадқиқот натижаларининг илмий ва амалий аҳамияти:

- “Компьютер лингвистикаси”, “Умумий тилшунослик” ва “Ҳозирги ўзбек тили” ҳамда амалий математика фанларини ўқитишда илмий-услубий манба сифатида фойдаланилади;

- олий ўқув юртиларидаги ўзбек филологияси ва амалий математика факультетларида битирув малакавий ишлари ҳамда магистрлик диссертациялари ёзишда истифода этилади;

- дарслик ва ўқув қўлланмалари ҳамда турли типдаги луғатлар тузишда муҳим ва қимматли манба(материал) ҳисобланади;

- тил ва компьютер алоқаси тарихи, масалалари ва истиқболларини илмий-лингвостатистик жиҳатдан кенгроқ ва асослироқ ўрганишда фойдаланилади ҳамда тадқиқот натижалари ва тавсиялари бошқа ёдномалар конкордансини, лингвостатистик параметрларини аниқлашда катта аҳамиятга эга.

Натижаларнинг жорий қилиниши. Тадқиқот натижалари СамДУ ва бошқа олий ўқув юртилари Умумий тилшунослик кафедраси профессор-ўқитувчилари СамДУ ўқув-услубий кенгаши томонидан маъқулланган ва нашрга тавсия этилган “Математик ва компьютер лингвистикаси” (2007) ва “Матни ўрганишининг лингвостатистик методлари” I қисм (2008) услубий қўлланмалари асосида бакалаврият йўналиши: 5220100 — филология (ўзбек филологияси) бўйича 30 соат маъруза ва 44 соатлик амалий машғулот (жами 74 соат) ни ҳам мунтазам ўтказиб келмоқдалар ва буларда мазкур иш материаллари муҳим манба сифатида фойдаланилмоқда;

- машина таржимаси йўналиши, информацион-статистик ҳамда жиёсий-типологик илмий изланишларни йўлга қўйиш имконияти яратилди;

- тилшунос, математик ва дастурчи (программист)ларнинг ҳамкорлигида ўзбек тили лингвистик банкларининг, машина фондининг асослари яратилишига йўл очилди.

Ишнинг синовдан ўтиши (апробацияси). Диссертациянинг асосий мазмуни “Ўзбек тилшунослигида лингвостатистик муаммолар” номли монография

(Т.: Фан, 2006. – 296 б.)да ўз ифодасини топган. Тадқиқот натижалари бўйича ЎзДЖТУ ва ТДИУ профессор-ўқитувчилари илмий-назарий анжуманлари (1996-2008)да, ҳар икки йилда мунтазам ўтказиб келинаётган Республика “Ўзбек тили” доимий анжуманларида, шунингдек, Ленинграддаги “Нутқ статистикаси” гуруҳининг ўрта осийлик (тошкентлик) аъзоси сифатида Москва (1976), Ленинград (1977), Кипинёв (1971, 1979), Махачқалъа (1978), Олма-Ота (1973,1976), Чимкент (1976, 1980, 1990, 1995) ва Тошкент (1980, 1991, 1995, 1996, 2001, 2004, 2006, 2007, 2008), Самарқанд (1989, 1991, 1999), Бухоро (1992), Урганч (1992, 2005), Термиз (2007), Қарши (1993, 2008) шаҳарларида ўтказилган Халқаро, Республика анжуманлари ва симпозиумларида маърузалар қилиб келинган.

Натижаларнинг эълон қилинганлиги. Тадқиқот ишининг асосий мазмуни ва натижалари олтига монография, саккизта рисола, ўн бешта журнал мақолалари ҳамда йигирма еттита тўшам мақолалари ва тезисларда акс этган.

Диссертациянинг тузилиши ва ҳажми. Кириш, беш боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловалардан иборат бўлиб, умумий ҳажми 301 саҳифадан ташкил топган.

2. ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Диссертациянинг Кириш қисмида мавзунинг долзарблиги, муаммоларнинг ўрганилиш даражаси, тадқиқотнинг илмий ишлар билан боғлиқлиги, тадқиқотнинг мақсади, олдига қўйган вазифалари, манбалари, методологик асослари ва усуллари, олинган натижалар ва уларнинг илмий янгилиги, ҳимояга олиб чиқилган асосий ҳолатлар, тадқиқотнинг илмий ва амалий аҳамияти, жамоватчилик синовидан ўтиши, тадқиқот натижаларининг эълон қилиниши, ишнинг тузилиши ва унинг ҳажми тўғрисида маълумот берилган.

Тадқиқотнинг “Лингвостатистик адабиётлар таҳлили ва статистик танланмаларни ташкил этиш хусусиятлари” номли биринчи бобида хориж лингвостатистикаси ва ўзбек тили материаллари асосидаги лингвостатистик тадқиқотлар, шунингдек, ҳарф ва фонемаларни биргаликда тадқиқ этилиш сабаблари, мақсад ва вазифалари ҳамда лингвостатистик танланмалар моҳияти(манбалари) тўғрисида фикр-мулоҳазалар баён этилди.

Тил ёки нутқ ҳодисаларини лингвостатистик тадқиқ этишга доир Д.Гильберт, В.Я.Буняковский, П.Б.Струве, Ю.И.Левин, А.А.Марков, В.А.Москович, А.А.Чупров, А.М.Пешковский, В.А.Богородицкий, В.Ф.Чистяков, Б.К.Крамаренко, Р.Г.Пиотровский, К.Б.Бектаев, А.А.Ляпидовский, Э.А.Штейнфельдт, Ю.Марков, И.И.Резвин, В.А.Артёмов, В.В.Шеворошкин, Е.В.Падучева, И.А.Мельчук, Р.М.Фрумкина, Л.Н.Засорина, А.Х.Джубанов, А.А.Ахабов, П.М.Алексеев, Г.В.Ермоленко, В.И.Перебийнос, В.М.Андрющенко, С.Г.Ахметова, В.А.Москович, Л.И.Ройзензон, В.Г.Адмони, Л.Зиндер, А.Н.Колмогоров, А.М.Кондратов, Б.В.Сухотин, А.Б.Долгопольский, А.А.Новак, А.В.Зубов, Б.Н.Головиц, Л.И.Ешан, А.Я.Шайкевич, Д.С.Лебедев, В.А.Гарман,

Ҳ.Арзикулов ва И.А.Киссен каби олимларнинг ишлари диссертацияда тўлиқ таҳлил этилган.

Ўзбек лингвостатистикасида Ўзбек тили матбуотининг лексик таркибини биринчи бўлиб текширган олим В.В.Решетов¹ эди. Унинг ишида Ўзбек тили матбуотидаги сўзларнинг келиб чиқиши (Ўзбекча, арабча, форсча, русча ва бошқалар) ва сўз туркумлари частотаси ўз ифодасини топган. 1940 йилда А.К.Боровков ҳам газета материаллари бўйича тадқиқот олиб борган ва натижаларини ўзбек орфографияси муаммоларига бағишланган мақоласида келтирган². Шунингдек, Ш. Шоабдурахмонов, Ғ.Абдурахмонов, И. Кўчқортоев, Б.Бафоев, А.Мамажонов, Ш.Сафаров, С.Каримов, Б. Иулдошев, С. Мухамедова, С. Отамирзаева, Д. Ўринбосва кабилар ҳам ўзларининг номзодлик ва докторлик диссертацияларида статистик методдан фойдаланганлар. И.А.Киссеннинг 1964 йилда нашр этилган “Қизил Ўзбекистон” газетасининг 1939-1940 йиллардаги бош мақолалари асосидаги ва ҳозирги ўзбек бадий прозасида сўзлар частотаси ҳақидаги мақола³да ўзбек бадий прозасидаги сўзлар лексик, грамматик ва семантик нуқтан назардан статистик тадқиқ этилган.

Рус ва ўзбек тилларининг қиёсий фонетикасида статистик методни тажриба сифатида дастлаб қўллаган В.М.Поповнинг “Ўзбек ва рус тилларини қиёсий ўрганишда статистик методни қўллаш” номи мақола⁴ характерлидир. Чунки қардош ва қардош бўлмаган тилларнинг товуш таркиби ёки морфологик хусусиятларини қиёсий ўрганиш орқали ўша тилларнинг ўзига хос хусусиятлари объектив равишда очилади. Бу ҳам илмий-назарий, ҳам амалий аҳамията молликдир.

Ўзбек тилида частотали лугат яратиш борасида Р.И.Битаев, Э.Р.Гукасянц, Ғ.Н.Михайлов, Ф.Н.Нигматова, И.Н.Соловьёва ва Ғ.Ш.Шариповлар томонидан тузилган “Русча-ўзбекча частотали лугат-минимуми”⁵ ва И.А.Киссеннинг 100.000 сўз (сўз — “словоупотребление” маъносида) ҳажмида бадий-прозаик асарлар асосида ҳозирги ўзбек тилининг частотали лугати чоп этилган. И.А.Киссеннинг бу лугати ҳам кўпроқ методик характерда яратилган. Проф. С.Каримов, А. Қаршиев, Ғ. Исроиловаларнинг ёзувчи Абдулла Қаҳҳор асарлари асосида 2007 йилда “Ўзбекистон Миллий энциклопедияси” Давлат илмий нашриёти томонидан чоп этилган “Алфавит-частотали”, “Частотали” ва “Терс” лугатлари эътиборга лойиқ.

¹ Решетов В.В. Лексический состав современной узбекской прессы // Проблемы языка. Вып 1. –Т., 1934. - С.41-51.

² Боровков А.К. Изменения в области узбекской лексики и новый алфавит (на основе русской графики) // Известия УзФАН СССР. – Т.: Фан, 1940. - №7. -С.18 - 37.

³ Киссен И.А. Опыт статистического исследования частотности лексикни передовых статей газеты “Қизил Ўзбекистон”. - “Научные труды ТашГУ им. В.И. Ленин”. Новая серия, вып. 247 (Ирасская, тюркская, арабская филология. Лингвостатистика). – Т., 1964. –С.44 - 58; Киссен И.А. Некоторые данные о частотности слов в современной узбекской художественной прозе. Юкоридаги тўқма. –С.59-73.

⁴ Попов В.М. Ўзбек ва рус тилларини қиёсий ўрганишда статистик методни қўллаш. - А.С.Пупкин номидаги Тил ва адабиёт институти асарлари. 1 китоб. “Ўзбек тили грамматикасида материаллар”. – Т.: ЎзССР ФА, 1949. -С.175-179.

⁵ Битаев Р.И., Гукасянц Э.Р., Михайлов Ғ.Н., Нигматова Ф.Н., Соловьёва И.Н., Шарипов Ғ.Ш. Частотный русско-узбекский словарь-минимум. – Т.: Уқитувчи, 1967.

Умуман, республикадаги рус ва ўзбек мактаблари учун ўзбек тили бўйича махсус ва алоҳид-алоҳида частотали лугатлар¹ яратилиши ҳам илмий, ҳам амалий аҳамиятта эга бўлган асосий вазифалардан ҳисобланади.

Ҳозирги ўзбек тилидаги ҳарф ва фонемаларни лингвостатистик тадқиқ этишда бирга олинишининг сабаби - ўзбек тилининг фонематик структурасига графиканинг қанчалик мос келиш-келмаслигини, уларнинг ўзаро муносабатини (қўлланиш частотаси нуқтан назаридан ҳам) аниқлаб беришдир.

Ўзбек тилининг график тизими унинг фонематик (фонологик) тизимига тўлиқ мос келмайди. Дувё тилларининг кўпчилигида ҳам худди шундай ҳолат мавжуд. Жаҳонда энг кўп тарқалган тиллардан бўлган - инглиз тили алифбосида ҳам 26 та ҳарф 44 та товушни ифодалайди.

Кирилл ёзуви (графикаси)га асосланган ўзбек алифбосидаги е,ё,ю,я ҳарфлари 2 товушнинг (ундош "й"+ унли) нинг ифодасидандир. Яъни "е = й + э; ё = й + о; ю = й + у; я = й + а" дан иборат. Шунингдек, ўзбек алифбосидаги "ж" ҳарфи (белгиси) ҳам сирғалувчи "ж" ни, ҳам қоршиқ (аффрикат) "дж" ни ифодалайди.

Ўзбек алифбоси сирасида 27-ўриндаги "ъ" (айриш белгиси) ва 28-ўриндаги "ь" (компатиш белгиси) товушни ифодаламайди (уларнинг функциялари бошқа-бошқадир).

Англашладикки, ҳарф бошқа-ю, фонема (товуш) бошқа. Умуман, у ёки бу товушнинг ёзувдаги шартли белгиси ҳарфдир. Товуш нутқнинг фонетик жиҳатдан бўлинмайдиган бирлигидир. Товушнинг сўз маъноларини фарқлашга хизмат қиладиган тури фонемадир. Ҳар қандай фонема - товуш, лекин ҳар қандай товуш - фонема эмас. Фонема - тил бирлиги, товуш эса нутқ бирлигидир.

Шуни ҳам алоҳида қайд этиш жоизки, ўзбек тилидаги икки ҳарф орқали ифодалаб келинаётган "нг" бурун товуши алифбода ўз ўрнига эга эмас. Кўпгина туркий тилларда, чунончи, қозоқ тили алифбосида мазкур товуш ўз ўрни ва ўз белгисига эга. Демак, ўзбек тилида бир ҳарф икки товушни ифодалайдиган, икки товуш бир ҳарф орқали ифодаланадиган, икки ҳарф бир фонема(товуш)ни ифодалайдиган ҳолатлар мавжуд. Ана шу жиҳатдан ҳам ўзбек ҳарфлар статистикаси билан фонемалар статистикаси бир-бирларидан маълум даражада фарқ қилади. Жумладан, ҳарфлар статистикасида қайси тартибда жойлаштирилган 31- ўриндаги "ж" нинг частотаси 850 бўлса, фонемалар статистикасидаги аффрикат "ж" (дж) частотаси 814 ва сирғалувчи "ж" нинг частотаси эса 36 дир. Тадқиқот жараёнида статистик кузатиш учун олиниши лозим бўлган манба(матн)ларда танланма усул(метод)лардан фойдаланилди.

Лингвостатистикада ҳам бошқа фанлардаги каби тадқиқ этиш методларидан бири танланма методидир. Чунки тилнинг ҳам бирор бир стили доирасида у ёки бу тил ҳодисасини лингвостатистик тадқиқотда, бутун бир стилга оид барча матнларни

¹ Ҳозирда кўпгина тиллар бўйича 1000 га яқин турли тилдаги частотали лугатлар мавжуд.

ёшасига, тугал кузатиш мумкин эмаслиги табиий бир ҳолдир. Шу сабабли ўша услубга (стилга) оид асарлардан танланмалар олинади. Жумладан, мазкур ишда ҳам тўрт хил функционал стиллар (проза, поэзия, ижтимоий-гуманитар: публицистика, фалсафа, тарих ва аниқ фанлар; математика, физика) нинг ҳар бирдан 50 000 дан (5000 дан ўн танланма) ҳарф олиниб, лингвостатистик таҳлил этилди.

Бунда олинган ҳар бир услубга доир манбалардаги ҳарфлар қўлланиш частотаси аниқлангач, улар олти гуруҳга ажратилиб, ўзаро солиштириб чиқилди. Статистик кузатишнинг барча гуруҳлари бўйича ҳарфлар частотаси тақсимланишининг таҳлили қуйидагича:

Танланмаларнинг ҳажмлари ва усуллари турлича бўлиши мумкин. Бу эса лингвостатистик тадқиқот олиб борувчининг мақсадига боғлиқдир.

Кузатиш учун тўпланган танланмаларнинг лингвостатистик таҳлили натижалари асосида бош тўплам (“генеральная совокупность”) бўйича фикр юргизиш мумкин. Бундай тадқиқот, умуман, статистикада, хусусан, лингвистик статистикада ҳам тўла тасдиқланган.

Лекин шуни унутмаслик керакки, олинган танланмалар ҳажми бош тўплам бўйича фикр-мулоҳаза юргизишга асос бўладиган даражада етарли бўлиши шарт, аксинча, ишнинг қиймати пасаяди.

Умуман, лингвостатистикада ҳам танланма ҳажми (N) ни қониқарли белгилаш имкони мавжуд.

Лингвостатистик тадқиқотларда танланма ҳажми қанчалик кўп ва хилма-хил бўлса, кутилган натижа шунчалик ишонарли ва самарали бўлади.

Бинобарин, ўзбек тили алифбоси ҳарфлари сатҳи асосидаги лингвостатистик тадқиқот учун олинган танланма ҳажми миқдорининг (N=200000) ҳам қониқарли эканлиги аниқланди. Чунки 100 % атрофидаги ҳарфлар (шунингдек, фонемалар) қўлланиш частотасининг нисбий хатоси¹ экспериментал аниқланди, яъни (аниқроғи 23%) $\delta = 0,2$ дан ошмади. Ўзбек алифбосидаги ҳарфларнинг (“А” билан биргаликда) 92 (91.66666) % (ёки 33 ҳарфининг) нисбий хатоси $\delta=0,11$ дан ошмаганлиги аниқланди.

Лингвостатистик тадқиқот учун олинган танланма ҳажми асосида ҳозирги ўзбек тилидаги ҳарф ёки фонемаларнинг стабиллик ва ностабиллик ҳолатларининг аниқ чегарасини белгилаш имкони мавжуд. Биринчи гуруҳдаги ҳарфлар умумий матн ҳажмининг қарийб ярми(50 %)ни, иккинчи гуруҳдагилар — 25 %, учинчи гуруҳдагилар — 15 %, тўртинчи гуруҳдагилар — 7 % ва бешинчи-олтинчи гуруҳдагилар — 3 % ташкил этади. I — IV гуруҳларда энг кўп қўлланидиган

¹ Нисбий хато $\delta = \frac{\sigma_p}{\sqrt{N \cdot P}}$ формула асосида аниқланди. Қараг: Фрумкина Р.М. Некоторые вопросы

методики составления частотных словарей. Сб. “Машинный перевод и прикладная лингвистика”. Вып. 2 (9). –М., 1959. –С. 27; математик-профессор, ф.ф.д. К.Б.Бектасовнинг курсатишича, нисбий хатоня қуйидагича аниқлашни

миьқулроқ, деб топди: $\delta = z \cdot \sqrt{\frac{1-P}{N \cdot P}}$. Умуман, тил брияклариня лингвостатистик тадқиқ этишда бу даражада

аниқликнинг талаб этиляши унчалик шарт эмас.

харфлар ҳам, энг кам қўлланадиган харфлар ҳам қарийб бир хил. I—II гуруҳдаги тўққизта ҳарфнинг қўлланиш частотаси деярли бир хил. Тафовутга олиб келувчи харфлар: Б, К, Е, Ф, Ш, М, Ц кабилардир.

Диссертациянинг “Ҳозирги ўзбек тилида ҳарфлар ва икки ҳарфли бирикмалар қўлланиш частотасининг лингвостатистик тадқиқи” номи иккинчи бобида “Ҳарфларнинг қўлланиш частотаси”(1-жадвал), “Ҳозирги ўзбек тили матнларида ҳарфлар қўлланиш частотасини статистик ишлаб чиқиш натижалари”, “Ҳозирги ўзбек тили матнлари энтропияси”, “Икки ҳарфли бирикмалар частотаси” каби масалалар тадқиқ этилган.

1-жадвал

Ўзбек тилида ҳарфлар қўлланиш частотаси
(камайиш тартибида) N = 200000¹.

i нинг қиймати	Ҳарф (белги)	F	P	i нинг қиймати	Ҳарф (белги)	F	P
1	2	3	4	5	6	7	8
1	“А”	27513	0,137565	20	З	2842	0,014210
2	А	25334	0,126670	21	Й	2460	0,012300
3	И	23271	0,116355	22	Ч	2183	0,010915
4	Н	13083	0,065415	23	В	2028	0,010140
5	Л	10539	0,052695	24	Ҳ	1922	0,009610
6	Р	10241	0,051205	25	П	1026	0,005130
7	О	7613	0,038065	26	Я	1020	0,005100
8	Т	7222	0,036110	27	Э	942	0,004710
9	Д	7001	0,035005	28	Х	919	0,004595
10	Г	6246	0,031230	29	Ё	870	0,004350
11	М	5976	0,029880	30	Ф	867	0,004335
12	Б	5937	0,029685	31	Ж	850	0,004250
13	У	5696	0,028480	32	Ғ	833	0,004165
14	К	5526	0,027630	33	Ю	347	0,001735
15	С	4756	0,023780	34	Ъ	317	0,001585
16	Қ	4265	0,021325	35	Ц	297	0,001485
17	Е	3582	0,017910	36	Ь	73	0,000365
18	Ў	3300	0,016500				
19	Ш	3103	0,015515		Σ	200000	1,000000

¹ Бунда N- танланма ҳажми; F - абсолют частота; P - нисбий частота. Шунингдек, камайиш тартибида жойлаштирилган ҳарфларнинг илганиш частотаси (“i нинг қиймати”). Базан бир европа тилларидаги энг кўп қўлланадиган тўрт ҳарф: инглиз тилида: E, T, A, O; немис тилида: E, N, I, S; француз тилида: E, S, I, A; латин тилида: A, I, S, U. Булар ҳақида қараг: Якубайтис Э.А. Оптимальный код алфавита латышского языка. - “Труды Института электроники и вычислительной техники”, Автоматика и вычислительная техника, П. - Рига. 1962. - С. 14; фин тилида: I, T, A, E; венгер тилида: E, A, T, I. Булар ҳақида қараг: Хасулнен Л. Развитие и структура финского языка, часть I (фонетика и морфология). - М.: ИЛ. 1953. - С. 15.; рус тилида: O, E, A, H (Белогов Г.Г., Фролов Г.Д. Эмпирические данные о распределении букв в русской письменной речи. - Сб. ст. “Проблемы кибернетики”, вып. 9. - М., 1963. -С. 287); ўзбек тилида эса жадвалда келтирилганидек А, И, Н, Л кабилар.

Ўзбек тили ҳарфлари билан боғлиқ ахборотларни автоматик қайта ишлаш учун ва оптимал кодировка (иккилик система асосида: 0,1) қилиш мақсадида ўзбек тили алифбоси учун Д.А.Хаффмен¹нинг иккилик коднинг тузиш, шунингдек, ҳар бир ҳарфнинг информатив қийматини белгилаш ҳамда ўзбек тили ҳарфлари билан боғлиқ ахборотларни узатиш тизимини жадаллаштириш мумкин.

Диссертациядаги 10-12- жадвалларда ҳозирги ўзбек тилининг босма матнларида сўзнинг боши, охири ва ўртасидаги ҳарф қўлланиш частотаси алоҳида-алоҳида қайд этилган. Шунингдек, диссертациянинг 13-24- жадвалларида ҳар бир функционал стиль бўйича, ҳозирги ўзбек тилидаги ҳарфлар қўлланиш частотасининг сўздаги ўрнига кўра тақсимланиши ҳамда олинган умумий танланма ҳажми доирасида ҳарфлар қўлланиш частотаси фоиз миқдорининг график тасвири келтирилган.

Ҳозирги ўзбек тилининг кўрсатилган ҳар бир стили бўйича ҳарфлар қўлланиш частотасининг ўртача частотасининг² - $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ аниқлаш мақсадида 50000 ҳарф 5000-5000 дан ўн танланмага ажратилди. Ўртача частотага статистик характеристика бериш учун дисперсия - $D(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, ўртача квадратик фарқ

- $\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$, “ҳақиқий ўртача частота”ни аниқлашдаги йўл қўйилиши

мумкин бўлган эҳтимол хато - $L = \frac{t \cdot \sigma}{\sqrt{n}}$ бунда “t” ўн танланма учун 2,26 га тенг

(тайёр жадвалдан олинди)³; “ҳақиқий ўртача частота” - $\overline{DX} = \bar{X} \pm L$, бунда $\overline{DX} = P(\bar{X} + L \leq \bar{X}_0 \leq \bar{X} - L) = 0,95$, вариация коэффициенти - $V = \frac{100 \cdot \sigma}{\bar{X}}$ ва нисбий хато

- $\delta = \frac{t \cdot \sigma}{\bar{X} \cdot \sqrt{n}}$ ҳам аниқланди (юқоридаги формулаларнинг ҳаммасида “n” - танланма сони, деб тушунилмоғи керак).

Диссертациядаги 25-28-жадвалларда ўзбек тилининг тўрт функционал стилилари босма матнларида ҳарфлар қўлланиш частотаси тақсимланишининг статистик ишлаб чиқиш натижалари келтирилди. Ўша жадваллардан кўринадики, танланма частоталарнинг ўртача частота атрофидаги тарқалиш (ёйилиш)

¹ Хаффмен Д.А. Метод построения кодов с минимальной избыточностью / “Кибернетический сборник”. Вып. 3. - Тбилиси: Иностр. лит-ра, 1961. - С. 79-87; Божучава Н.В. Двоичный код Хаффмена для алфавита грузинского языка / “Труды Тбилисского гос. университета”. Серия кибернетических наук. Т.120. - Тбилиси, 1967. - С.193-195; Кечухашвили Г.Н. Об информативной ценности букв русского алфавита / “Труды Тбилисского гос. университета”. Серия кибернетических наук. Т. 120. - Тбилиси, 1967. - С.145-164.

² \bar{X} , шунингдек, $D(X)$, σ , L , V , δ каби қўлланилган формулалар ҳақида қаранг: Ван дер Варден Б.Л. Математическая статистика. - М., 1960; Гутер Р.С. и Олчвинский Б.В. Основы теории вероятностей. - М.: Просвещение, 1967; Soatov N.M., Tillaov Jajueva G.N. Statistika. - Т., 2005.

³ “t”нинг қиймати танланма сонига қараб белгиланади, яъни “t” танланма сонига боғлиқдир (“t” - узгармас коэффицент - константа).

даражаси турличадир. Барча матнлар бўйича шуни қатъий айтиш мумкинки, каерда дисперсия “Д(X)”нинг қиймати катта бўлса, ўша ерда ўртача квадратик фарқ “σ” ва “ҳақиқий ўртача частота”ни аниқлашдаги йўл қўйилиши мумкин бўлган эҳтимол хато (L) нинг миқдори ҳам катта бўлиши билан характерланади.

Хусусан, ўзбек тилининг проза матнларидаги – а, л, м, р, “Δ”, т ҳарф (белги)ларида кўрсатилган учала [Д(X), σ, L] ҳолат нисбатан каттадир. Шунингдек, поэтик матнлар¹ бўйича - и, н, д, б, а, к, “Δ”, р; ижтимоий-гуманитар матнларидаги - и, н, л, “Δ”, а, р, т, м, д, с; аниқ фанлар матнларидаги - а, н, р, и, л, м, о, у, с, к, б хаби ҳарф (белги)ларида ҳам юқоридаги ҳолат қийматларини кўриш мумкин. Лекин аниқ фанлар матнларида “Д(X)”, “σ”, “L”ларнинг бошқа матнларга нисбатан энг катта қийматга эгаллиги танланма частоталарнинг ўртача частотадан тарқалиш даражаси олинган матнларнинг тўла ҳолда бир хил эмаслигини яққол тасдиқлайди.

Математик статистикада энг кўп қўлланадиган тушунчалардан бири - ўртача квадратик фарқнинг ўртача частотага бўлган фоизли нисбати - вариация коэффициенти ҳисобланади. Диссертацияда вариация коэффициенти нуқтан назаридан ҳозирги ўзбек тилининг тўртгала стиллари матнлари бўйича фоиз нисбатларидаги ҳарфларнинг ўзаро фарқи келтирилган. Айниқса, ҳозирги ўзбек тилининг аниқ фанлар учун танланган матнлари бошқа стилдаги матнлардан ўзгачалиги билан фарқланади.

Маълумки, тил ва нутқ бирликларини лингвостатистик тадқиқ этишда, айниқса, ўртача частотага нисбатан берилаётган тавсифда нисбий хато (δ) миқдорининг кўп ҳолларда 10 фоиздан ошмаслиги мақсадга мувофиқ, деб қаралади. Диссертацияда нисбий хато миқдорларининг фоиз ифодалари алоҳида-алоҳида стиллар бўйича қайд қилинган.

Информация миқдорининг, энтропиянинг, код оптималлигининг статистик методда аниқланиши тилининг фонологик системасини тадқиқ этишда энг фойдали - эффектив натижалар беради. Шунингдек, ҳарф(мати) энтропияси аниқ баҳосининг белгиланиши математик лингвистиканинг долзарб масалаларидан бири ҳисобланади.

Информация миқдори ўлчови - энтропия ҳисобланиб, у информация назариясида “бит”да ўлчанади. “Бит” - иккилик рақамнинг (0, 1) энг оддий ва қулай қисқартирилган ифодасидир. Умуман, вақт бирлиги сифатида “сония”, оғирлик бирлиги сифатида “грамм” қабул қилинганидек, информация назариясида информация миқдори бирлиги сифатида “бит” ёки “ҳа-йўқ” бирлиги (0, 1) қабул

¹ Танланган поэтик матнларда “и” босан “ц”нинг ҳатто танланма доирасида ҳам жуда кам қўлланилиши туфайли ўртача частота ва унга нисбатан ҳеч қандай таъсиф берилмади.

қилинган. Бинобарин, у ёки бу хабарда (мас.: ҳарф, фонема, бўғин, морфема ва сўзшакл кабиларда) қанча “бит” бўлса, энтропия ҳам шунга тенг бўлади.

Информация назариясида энтропия у ёки бу символга (харфга) ўртача иккилик белги миқдорини беради. Бу эса кодировка системасини энг қулай ташкил этишда муҳим манба бўлиб хизмат қилади. Шунингдек, ўзбек тилининг босма матнлари энтропиясининг аниқланиши шу тил ҳарфлари билан боғлиқ информацияларнинг узатиш тезлигини жадаллаштиришда аҳамиятлидир.

Бир ҳарф қанча информация беришининг мумкин қадар тўғри ҳал этилиши компьютер лингвистикасида ўзбек тили матнларини таҳлил қилиш учун жуда зарурдир.

Маълумки, ҳозирги ўзбек тилининг алифбосида 35 шакл мавжуд бўлиб, пробел (“Δ”) тизим 36 тадир. Демак, пробелни ҳам ҳисобга олганда К.Шеннон формуласи¹ асосида ўзбек тили алифбосининг энтропияси қуйидагичадир:

$$H_0 = -\sum_{i=1}^n P_i \log_2 P_i = -\sum_{i=1}^{36} P_i \log_2 P_i = -36 \cdot \frac{1}{36} \log_2 \frac{1}{36} = -\log_2 \frac{1}{36} = \log_2 36 = 5,17$$

иккилик белги.

Англашиладики, ўзбек тилининг бир ҳарфи натижада 5,17 “бит”да ёки $36=2^{5,17}$ информация беради(2-жадвал).

2-жадвал

Ҳарфлар энтропияси

ҳарф (белги)	- P log ₂ P	ҳарф (белги)	- P log ₂ P	ҳарф (белги)	- P log ₂ P	ҳарф (белги)	- P log ₂ P
“Δ”	0,393593	м	0,151043	й	0,078536	ж	0,033161
а	0,377323	б	0,150316	ч	0,071570	ғ	0,033161
и	0,361336	у	0,145918	в	0,067476	ю	0,015640
н	0,257570	к	0,142946	х	0,064346	ъ	0,014860
л	0,223487	с	0,128351	п	0,038838	ц	0,014071
р	0,219531	қ	0,118690	я	0,038838	ь	0,004168
о	0,179279	е	0,104325	э	0,036346	Σ = 4,345658 ≅ ≅ 4,35	
т	0,173321	ё	0,098150	х	0,035715		
д	0,169278	ш	0,093636	ё	0,033804		
г	0,156072	з	0,087159	ф	0,033804		

Ўттиз олти символдан бирини танлашда ≈ 5,17 иккилик белги мувофиқ келади. Агарда коднинг эффективлиги аҳамиятсиз (бефарқ) бўлса, у ҳолда ҳар бир

¹ Пирс Дж. Р. Символы, сигналы, шумы (закономерности и процессы передачи информации). - М.: Мир, 1967. Гл. V. “Энтропия”. - С. 96 - 128; Shannon C.E., Prediction and entropy of printed English, “BSTJ”, vol. XXX, № 1, 1951. - pp. 50-64; Хвичин А.Я. Понятие энтропии в теории вероятностей // Успехи математических наук. Т. VIII. Вып. 3 (55). - М., 1953. - С. 3-20.

символ (харф) 5 белгили иккилик сонни акс эттириши мумкин, бунда 5 сони фойдаланилмай қолади¹.

Ҳозирги ўзбек тилининг матнларидаги иккиҳарфли бирикма (биграмма)лар қўллаганин частотасини статистик тадқиқ этиш бирмунча мураккаб масалалар сирасига киради. Бундай тадқиқотларда ЭҲМ (компьютер)ларни қўллаш туфайлигина самарали натижаларга эришиш мумкин.

Диссертациянинг учинчи бобида “Ҳозирги ўзбек тили фонологик структурасининг лингвостатистик тадқиқи” баён қилинган. Статистик тадқиқ тилининг фонематик таркиби жиҳатидан нутқий оқимда қай даражада оҳангдорлиги (сонантлилиги)ни аниқлашга ёрдам беради. Ҳозирги ўзбек тили фонологик системасидаги фонемалар миқдори масаласида (айниқса, унлилар системасида) турлича қарашлар мавжуд². Ҳозирги ўзбек тилининг фонематик таркиби кўпчилик тилшунослар томонидан қабул қилинган мавжуд дарслик, қўлланма ва монографияларда кўрсатилган олти унли ва йигирма беш ундош фонемалар асосида лингвостатистик таҳлил қилинди.

Ҳозирги ўзбек тилидаги фонемаларнинг қўллаганин частотасини статистик тадқиқ этиш учун турли мавзуларга оид матнлар (проза бўйича $N_1 = 42960$; поэзия бўйича $N_2 = 42603$; ижтимоий-гуманитар - тарих, философия, публицистика бўйича $N_3 = 43524$; аниқ фанлар - физика, математика бўйича $N_4 = 43389$) танланди ($\Sigma = 172476$).

Юқоридаги танланма ҳажмларига икки сўз оралиги - “пробел” ҳисобга кирмайди³; “пробел” билан эса $N_1 = 50118$; $N_2 = 49956$; $N_3 = 50150$; $N_4 = 49776$; жами $\Sigma = 200000$. Бу танланма ҳажми экспериментал йўл билан нисбий хато (δ) формуласи⁴ асосида текширилганлиги учун тўла равишда қониқарли баҳоланди, яъни ўзбек тилидаги барча фонемалар (“Δ” билан) 94 % нинг нисбий хатоси 0,1 дан ошмаганлиги аниқланди. Ўзбек тилида энг кам қўлланган “ц” да $\delta = 0,1$ (аниқроғи 0,11596), “ж”да эса $\delta = 0,3$ (аниқроғи 0,33333). “Пробел”сиз текширишлар ўзбек тилидаги барча фонемаларнинг 97 % ини ташкил этган ўттиз фонеманинг нисбий хатоси $\delta = 0,00362 + 0,10776$ ва энг кам қўлланган “ж” фонемасининг нисбий хатоси 0,3 дан ҳам ошмаганлигини кўрсатди.

Фонологостатистик тадқиқот учун олинган танланма ҳажмининг “ишончлилиқ баҳоси” тўла равишда қониқарлидир. Чунки барча лингвостатистик

¹ Пирс Дж. Р., ўша асар. -Б.92 (инглиз тили алифбосидаги 26 харф, пробел билан 27 символ ўрнида ўзбек тилининг фойдаланилиши).

² И.А.Киссеннинг кўрсатишича, ўзбек тилидаги фонемалар миқдори 36 тагача боради. Қаранг: Киссен И.А. Некоторые данные о частотности фонем в современном узбекском литературном языке / “Научные труды ТашТ У пн. В.И. Ленинга”. Вып. 229. Востоковедение (языки, литература, история стран Азии). - Т., 1964. - С.67-79; Киссен И.А. Курс сопоставительной грамматики русского и узбекского языков. - Т.: Учитель, 1966. - С.18; 23-24.

³ “Пробел”ни ҳисобга олишимиздан мақсад ўзбек тилида сўз узунлигининг ўртача миқдори неча фонема (товуш)дан иборатлигини аниқлашдир (“Δ” - рус тилидаги “нулевая фонема” маъносидан).

⁴ Бу ҳақда қаранг: Фрумкина Р.М. Некоторые вопросы методики составления частотных словарей (статистическая структура словаря в текста) / “Машинный перевод и прикладная лингвистика” (Бюллетень Объединения по машинному переводу). Вып. 2 (9). - М., 1959. - С.23-36.

тадқиқотларда нисбий хатонинг 0,3 дан ошмаслиги шарт қилиб олинган (“энг ишонччилик” ҳолатда $\delta=0,1$ дан ошмаслиги керак). Шу аниқликларда олиб борилган лингвостатистик тадқиқотлар натижалари амалда тасдиқланган¹. Ҳозирги ўзбек тили матида фонемалар қўлланиш частотаси ишлаб чиқилди.

Фонемаларни унли ва ундошлар системасига ажратиб, ўзаро физик нисбатлари солиштирилганда, унлилар системаси (ўзбек тилидаги барча фонемаларнинг 19 % и) умумий матн ҳажмининг 42 % и (аниқроғи 41,751896; $F = 72012$)ни, ундошлар эса 58 % и (аниқроғи 58,248104; $F = 100464$)ни ташкил этишидан кўринадики, ҳозирги ўзбек тилида фонологик нуқтаи назардан ўртача 100 унлига 140 ундош тўғри келади, яъни ўзбек тили учун унли ва ундошларнинг ўзаро миқдорий муносабати жиҳатидан $N_c: N_v = 1,40$ (1,3951008) статистик параметр демакдир.

Ҳозирги ўзбек тили фонематик тузилишининг характерли хусусиятларидан бири бошқа кўшгина тиллардаги каби ундошларнинг унлиларга нисбатан матнда (нутқда) кўп қўлланишидир (бундан гавай, самоа, маори ва фин тиллари мустаснодир).

Ҳозирги ўзбек тилида энг кўп қўлланадиган фонема “а” бўлиб, у барча унлилар қўлланиш частотасининг 37% ини ташкил этади. Ундан кейин “и” (32 %), “о” (12 %), “у” (8 %), “е (э)” (6 %), “ў” (5 %). Кўринадики, унлилар системасидаги икки унли фонема (а,и) барча унлиларнинг 2/3 қисмидан кўпроғини ташкил этади (69 %), яъни а, и > о, у, е (э), ў. Ўзбек тилидаги энг кўп қўлланадиган ундош - “н” фонемаси барча ундошлар қўлланиш частотасининг 11 % ини ташкил этади; ундан кейин “л” (10 %), “р” (10 %), “т” (7 %), “д” (7 %) ва ҳ.к.

Ўзбек тилидаги сонор ундошлар (в, л, р, м, й², ў)нинг умумий матндаги салмоғи 26 % (аниқроғи 26,024491)дир.

Тилдаги олти унли ва олти сонор (жамми 12 сонант)нинг умумий матндаги салмоғи 68 % (67,776387)дир³, яъни ўзбек тилини фонематик таркиби жиҳатидан тўла равишда сонантла, деб тавсифлаш мумкин.

Ҳозирги ўзбек тилидаги фонемаларнинг матнда қўлланиш частотаси фоизлари ҳисобига кўра (“Δ”сиз) қуйидагича (фонематик) формуласини тузиш мумкин:

$$A_{15}I_{13}(NLP)_6 O_5(TDU)_4 [MBK\dot{I}CE(\dot{E})]_3 (K\dot{G}\dot{U}S\dot{Z})_2 \\ (H\dot{C}V\dot{X}\dot{I}P\dot{X}\dot{F})_1 (F\dot{J}\dot{D}\dot{J})_{0,5}$$

¹ Перебийнос В.И. Частота и сочетаемость фонем современного украинского языка / “Семантар: Автоматизация информационных работ и вопросы прикладной лингвистики”. - Киев, 1965.

² Ўзбек тилидаги кўпчилик адабиётларда “й” фонемаси классификация бўйича шовқонилар туркумида, априлларда эса сонорлар туркумида кўрсатилган. (Sjoberg A.F. The Phonology of standard Uzbek. In: “American studies in Altaic Linguistics”, Indiana University Publications Ural and Altaic series., vol. XIII. 1962 – Bloomington, pp. 237-261.) Умуман, “й” фонемаси кўпчилик интервиккал холатда сонорланиб, ёқининг лингвостатистик кузатиш материалларида шу ҳолат кузлуклар учраганини учун “й” ундош фонемаси сонорлар туркумида ҳисобланади; “й”ни сонорлар туркумида ҳисобга олишмас, у ҳолда ўзбек тилидаги 5 сонор ундош умумий матн ҳажмининг 23 фоизи (23,111042)ни; шовқонилар эса “й” билан бирга (жамми 20 та) 35 фоизи (35,137062)ни ташкил этади.

³ “й”сиз 65 %.

Юқоридаги формулага кирган дастлабки олтига (3 унли: А, И, О ва 3 сонор: Н, Л, Р) фонема (Ўзбек тилидаги ҳамма фонемаларнинг 19 % и) умумий матн ҳажмининг ≈ 52 % ини; шунингдек, дастлабки 15 та (5 унли: А, И, О, У, Е (Э)), 5 сонор: Н, Л, Р, М, Ў ва 5 шовқинли: Т, Д, Б, Қ, С) фонема (ҳамма фонеманинг 48 % и) эса умумий матн ҳажмининг 82 % ини, яъни 4/5 қисмидан кўпроғини ташкил этади. Қолган 16 та фонема - Қ, Г, Ў, Ш, З, Ҳ, Ч, В, Ҳ, П, Х, Ф, Ф, Ж, Ц, Ж (ҳамма фонеманинг 52 % и) умумий матн ҳажмининг 18 % ини ташкил этади.

Ўзбек адабий тилидаги шовқинчилар (Б, П, В, Ф, Д, Т, Ж, Ч, Ж, Ш, З, С, Г, К, Ф, Х, Қ, Ҳ, Ц)нинг умумий матндаги салмоғи 32 % дир¹. Шундан жарангли ундош (Д, Б, Г, З, В, Ф, Ж, Ж)лар 14 %²; жарангсиз ундош (Т, К, С, Қ, Ш, Ч, Ҳ, П, Х, Ф, Ц)лар 19 % (18,614762); жуфт ундош (Д - Т, Б - П, Г - Қ, З - С, В - Ф, Ф - Х, Ж - Ч, Ж - Ш)лардан жарангли (F=23472)лари 14 %, жарангсиз (F=25617)лари 15 %. Кўринадикки, жуфт ундошларнинг Ўзбек тилидаги барча фонемаларга нисбатан муносабати - 14 : 15 ёки 48:52.

Шовқинли ундошлардан портловчилар (Т, Д, Б, К, Қ, Г, Ч, П, Ж, Ц)нинг умумий матндаги салмоғи 22 % бўлиб, шундан оддий портловчилар (Т, Д, Б, К, Қ, Г, П) умумий матн ҳажмининг 20 % ини; коришиқ (аффрикатив) портловчилар (Ч, Ж, Ц) эса 2 % ини ташкил этади. Сирғалувчилар (С, Ш, З, В, Ҳ, Х, Ф, Ф, Ж)нинг матндаги салмоғи 10 %³.

Ўзбек тилидаги лаб ундош (М, Б, В, П, Ф)лари барча ундошлар қўлланиш частотасининг 16 %ини ташкил этади⁴.

Лаб ундошларининг барча фонемаларга нисбатан салмоғи 9 %дир. Тил товушлари умумий матн ҳажмининг 48 %ини ташкил этади. Шундан, тил олди ундошлари (Н, Л, Р, Т, Д, С, Ш, З, Ч, Ж, Ц, Ж)нинг матндаги салмоғи 35 %; саёз тил орқа ундошлари (Қ, Г, Ҳ) 7 %; чуқур тил орқа ундошлари (Қ, Х, Ф) 3 %; тил ўрта ундоши (Й) эса 3 %дир.

Маълумки, Ўзбек тилидаги унлилар системаси уч артикуляцион параметр бўйича фаркланади. Тил орқа унлилари (У, Ў, О)нинг тил олди унлилари [И, Е (Э), А] га нисбатан матнда қўлланиш миқдорий муносабати 25 : 75; Ўзбек тилидаги ҳамма фонемаларга нисбатан эса 10:31дир (лаб иштироки ҳам шу нисбатдадир). Кенг унли (А, О)лар Ўзбек тилидаги барча фонемаларга нисбатан

¹ "к" билан бирга 35 %.

² "ж" билан бирга 17 %.

³ "г" билан бирга 13 %.

⁴ Солиштириб кўринг: туркий тиллардан - турк, вўгай ва татар тилларида 15 % дан; қорақалпоқ тилида 17 %; серб-хорват, чех, поляк тилларида, қадимги славян тилларидагидек 22 % дан; словен тилида 23 %; рус ва болгар тилларида 20 % дан (Қаранг: Никонов В.А. Интерпретация фонетических частот. - В кн.: "Ученые записки института славяноведения". Т. XXVII. - М., 1963. - С.259).

умумий матннинг 20 %ини, тор унли (И, У)лар 17 %ини; ўрта тор унли [Е (Э), Ў] лар эса 5 %ини ташкил этади.

Юқорида келтирилган фонематик формула асосида ҳозирги ўзбек адабий тилидаги фонемаларнинг “нутқий оқим”да жойлашиш тартибини¹ куйидагича келтириш мумкин (биринчиси “Δ” билан, иккинчиси “Δ” сиз):

Фонемаларнинг жойлашиш тартиби

	I	II
1	Δ	А
2	А	И
3	И	Н Л Р
4	Н Л Р	О
5	О Т Д	Т Д У
6	У М Б К Й	М Б К Й С Е (Э)
7	С Е (Э) Қ Г Ў Ш	Қ Г Ў Ш З
8	З Ҳ Ч В Ҳ П	Ҳ Ч В Ҳ П Х Ф
9	Х Ф Ҳ Ж Ц Ж	Ҳ Ж Ц Ж

Юқоридагидек тартибда жойлашган фонемаларни унлилар системасини алоҳида, ундошлар системасини алоҳида гуруҳлаштириб жойлаштириш фақат шу ўриндагина эмас, ҳатто тилининг ички функционал стиллари доирасида ҳам мумкин эмас. Чунки қаторлар бўйича унлиларнинг статистик тақсимланиши турличадир. Диссертацияда ҳозирги ўзбек тилининг турли стилларида фонемаларнинг қўлланиш частотаси ва уларнинг ўзаро миқдори муносабатлари алоҳида-алоҳида жадалларда қайд этилган.

Ҳозирги ўзбек тилининг турли стиллари бўйича фонемалар “нутқий оқим” (мат)да куйидагича жойлашади (биринчиси “Δ” билан бирга, иккинчиси “Δ” сиз):

	Проза матнлари бўйича		Поэзия матнлари бўйича	
	I	II	I	II
1	Δ	А	Δ	А
2	А	И	А	И
3	И	Н Р	И	Р
4	Н Р О	О Л Д	Р	Н Л О
5	Л Д	Б Т	Н Л О	Д У Т Б
6	Б Т У М К Й	У М К Й Қ Г Е (Э)	Д	М Й Қ Қ С
7	Қ Г Е (Э) С Ў Ш	С Ў Ш З	У Т Б М Й К	Ў Е (Э) Г З Ш Ҳ
8	З Ч Ҳ Ҳ В П Х	Ч Ҳ Ҳ В П Х Ҳ Ж	Қ С Ў Е (Э) Г З	Ч Ҳ В Ҳ П Х
9	Ҳ Ж Ф Ц Ж	Ф Ц Ж	Ш Ҳ Ч Ҳ В Ҳ П Х	Ж Ф Ж Ц
10			Ж Ф Ж Ц	

¹ Бу тартибни дастлаб В.Ф.Чистяков, Б.К.Крамаренко қўлаган. Улар проф. В.А.Богородицкий маълумоти бўйича унли ва ундошларни алоҳида тақсирган. Шунинг учун юқоридаги тартибни ундолларга кўра кўрсатган, холос. Тугал охиригача етказмаганлар Қаранг: Чистяков В.Ф., Крамаренко Б.К. Опыт приложения статистического метода к языкознанию. Вып. I. – Краснодар, 1929. - С. 14.

	Ижтимоий-гуманитар матнлари буйича		Аниқ фанлар матнлари буйича	
	I	II	I	II
1	A	A	A Δ И	А И
2	Δ	И	Н Л	Н Л
3	И	Л	Р	Р
4	Л	Р Н	Т	Т К О
5	Р Н	О	К О Д У М Е (Э) Б С	Д У М Е (Э) Б С Г
6	О Т	Т Д М	Г Й Қ Ү	Й Қ Ү Ч Ш Ң
7	Д М У Й К Б	У Й К Б С	Ч Ш Ң З В Х Ф	З В Х Ф
8	С Е (Э) Қ Ш Г	Е (Э) Қ Ш Г В	П Ж Ц Х Ф Ж	П Ж Ц Х Ф Ж
9	В З Х Ү Ң Ч Х	З Х Ү Ң Ч Х П Ж		
10	П Ж Ф Г Ц Ж	Ф Г Ц Ж		

Хозирги ўзбек тили стиллари буйича фонемаларнинг матнда қўлланиш частотасининг фонзлари ҳисобига кўра (“Δ” сиз) куйидагича фонематик формуласини келтирамыз:

Проза матнлари буйича

A₁₅ И₁₄ (НР)₆ (ОЛД)₅ (БТ)₄ [УМҚЙКГЕ(Э)]₃ (СҮШЗ)₂ (ЧҲНВПХФЖ)₁ (ФГЦЖ)_{0,5}

Поэзия матнлари буйича

A₁₅ И₁₂ Р₇ (НЛО)₆ (ДУТБ)₄ (МЙККС)₃ [ҮЕ(Э)ГЗШЦ]₂ (ЧХВФГХ)₁ (ЖФЖЦ)_{0,5}

Ижтимоий-гуманитар матнлари буйича

A₁₆ И₁₄ Л₇ (РН)₆ О₅ (ТДМ)₄ (УЙКБС)₃ [Е(Э)ҚШГВ]₂ (ЗХҮҢЧХПЖ)₁ (ФГЦЖ)_{0,5}

Аниқ фанлар матнлари буйича

(АИ)₁₅ (НЛ)₇ Р₆ (ТКО)₄ [ДУМЕ(Э)БСГ]₃ [ЙҚҮЧШҢ]₂ (ЗВХФ)₁ (ПЖЦХФЖ)_{0,5}

Проза матнлари буйича қўлланган дастлабки 7 фонема (А, И, Н, Р, О, Л, Д - ҳамма фонеманинг 23 %) умумий матн ҳажмининг 56 %ни (55,609869)ни; дастлабки 16 фонема (А, И, Н, Р, О, Л, Д, Б, Т, У, М, К, Й, Қ, Г, Е(Э) ҳамма фонеманинг 52 %ни) эса умумий матннинг 4/5 қисмидан кўпроғи (84,355212 %)ни ташкил этади. Қолган 15 фонеманинг матндаги салмоғи 15,644776 %дир.

Поэзия матнлари буйича қўлланган дастлабки 6 фонема (А, И, Р, Н, Л, О - ҳамма фонеманинг 19 %) умумий матн ҳажмининг ярмини (50,484708%); дастлабки 15 фонема (А, И, Р, Н, Л, О, Д, У, Т, Б, М, Й, К, Қ, С - ҳамма фонеманинг 48 % и) эса умумий матн ҳажмининг 4/5 қисминя (81,632749 %) ташкил этади. Қолган 16 фонеманинг матндаги салмоғи 18,367251 % дир.

Ижтимоий-гуманитар матнлари буйича қўлланган дастлабки 6 фонема (А, И, Л, Р, Н, О)нинг умумий матндаги салмоғи 54 % (53,540576); дастлабки 14 фонема (А, И, Л, Р, Н, О, Т, Д, М, У, Й, К, Б, С - ҳамма фонеманинг 45 % и) ники эса 81 % дир. Қолган 17 фонсема (ҳамма фонеманинг 53 % и) умумий матн ҳажмининг 19 % ини ташкил этади.

Аниқ фанлар матнлари буйича қўлланган дастлабки 5 фонема (А, И, Н, Л, Р - ҳамма фонеманинг 16 % и) умумий матн ҳажмининг 49,378875 % ини; дастлабки

15 фонема (А, И, Н, Л, Р, Т, К, О, Д, У, М, Е(Э), Б, С, Г) эса 83 % ини ташкил этади. Қолган 16 фонеманинг матндаги салмоғи 17 % дир.

Ҳозирги ўзбек тилидаги 6 унли фонема проза матнларида қўлланган ҳамма фонеманинг 42 % ини, 25 ундош фонема эса 58% ини ташкил этади. Частоталар муносабати ($\Sigma = 42960$ га нисбатан) $V_{18150} : C_{24810}$. Бундай матнлардаги “фонематик занжир”да ўртга 100 унлига ≈ 137 ундош қўлланади, яъни консонантли коэффицент - $Nc : Nv \approx 1,37$ дир.

Поэзия матнлари бўйича барча унлиларнинг умумий матндаги салмоғи 40 %; ундошларники 60 % дир. Частоталар муносабати ($\Sigma = 42603$ га нисбатан) $V_{17243} : C_{25360}$.

Поэзия матнларидаги “фонематик занжир”да унли ва ундошларнинг нисбий частота микдорий муносабати (консонантли коэффицентлари) - $Nc : Nv \approx 1,47$ дир.

Ижтимоий-гуманитар матнларида ўзбек тилидаги барча унлиларнинг умумий салмоғи 42 %; ундошларники 58 % дир. Частоталар муносабати ($\Sigma = 43524$ га нисбатан) $V_{18344} : C_{25180}$.

Бундай матнлардаги консонантли коэффицент - $Nc : Nv \approx 1,37$.

Аниқ фанлар матнларидаги барча унлиларнинг умумий матндаги салмоғи 42 %; ундошларники 58 %дир. Частоталар муносабати ($\Sigma = 43389$ га нисбатан) $V_{18275} : C_{25114}$. Бундай матнлардаги консонантли коэффицент - $Nc:Nv \approx 1,37$.

Фонологостатистик нуктаи назардан ҳозирги ўзбек тилининг барча прозаик матнларидаги консонантли коэффицент қиймати поэтик матнлардагидан фарқли ўларок “бир хилрок” ($Nc : Nv \approx “1,37”$)дир.

Ҳозирги ўзбек тили унлилар системасидаги икки унли фонема (А,И)нинг барча унлиларга нисбатан матндаги салмоғи қарийб 2/3 қисмига тенг:

[А, И > О, У, Е(Э), Ў].

Турли стиллар бўйича ҳозирги ўзбек тилидаги ҳар бир унли фонеманинг сўздаги ўрнига кўра қўлланишининг жойлаштирилиши диссертацияда келтирилган.

Ҳозирги ўзбек тилидаги олти унли фонема турли стиллар доирасида, шунингдек, сўзнинг турли позицияларидаги қўлланиш микдори жиҳатидан бир-биридан у ёки бу даражада фарқланади. Фонемаларнинг сўз боши, сўз ўртаси ёки сўз охиридаги қўлланиши турли стилларда турлича қийматга эга бўлиб, улардан баъзи стиллардаги қиймат бир-бирига яқин, баъзиларида эса фарқ каттарок эканлигининг қай даражадалигини тажриба тариқасида аниқ методлар ёрдамида таҳлил қилиш мумкин. Масалан, “о” фонемасининг сўз бошида келиши наср(проза)да 23,6 %, назм(поэзия)да эса 29,7 % бўлиб, бу иккала стилдаги фарқнинг даражасини аниқлаш учун математик статистикадаги аниқ формулалардан фойдаланилди. “О” фонемасининг прозадаги нисбий частотаси (P_1) ва таъланма ҳажми (N_1), шунингдек, поэзиядаги P_2 ва N_2 ларнинг қийматлари фойдаланилган формулалар бўйича куйидагича ҳисобланди:

$$t_{\rho_1 - \rho_2} = \frac{(\rho_1 - \rho_2)}{\delta_{\rho_1 - \rho_2}} = 4,209$$

Беш фонизлик қиймат даражада $t_{0,05} = 1,96 \approx 2$. Агар олинган натижа ($t_{\rho_1 - \rho_2}$) $t_{0,05}$ - қийматидан кичик бўлса, яъни $t_{\rho_1 - \rho_2} < t_{0,05}$, у ҳолда фарқ аҳамиятсиз; $t_{\rho_1 - \rho_2} \geq t_{0,05}$ бўлса, бу ўринда солиштирилаётган иккала стилдаги фонематик даражадаги фарқ аҳамиятлидир.

Демак, $t_{\rho_1 - \rho_2} \geq t_{0,05} = 1,96 \approx 2$; $t_{\rho_1 - \rho_2} = 4,209 > 1,96 = t_{0,05}$ экан.

Юқоридagi усулга биноан “а” фонемасининг турли стилларда қўлланиш миқдорининг ўзаро ўхшаш ёки фарқли томонлари қуйидача:

“а” фонемасининг проза матнларида барча унли фонемаларга нисбатан қўлланиш салмоғи 36,7 %, поэзия стилида эса 37,0 %. Демак, буларнинг матндаги салмоғи бир-бирига бирмунча яқин. Бироқ маълум даражада фарқ ҳам мавжуд. Мазкур фарқни эътиборга олиш керакми ёки йўқми деган саволни аниқлашда ҳам юқоридagi формулалар анча қўл келади. Натижа шуни кўрсатдики,

$$t_{\rho_1 - \rho_2} = 0,49958 \approx 0,5.$$

Демак, $t_{\rho_1 - \rho_2} = 0,5 < 1,96 = t_{0,05}$.

Шунга кўра, сўз ўртасида “а” (фонематик даражада) бадиий-прозаик услуб билан поэзия услуби матнлари орасидаги фарқ аҳамиятсиздир.

Юқорида фақат тажриба тариқасида “о” ва “а” фонемаларининг қўлланиш салмоғига доир икки стилини бир-бирига қиёслаб кўрдик.

Шу усул билан қолган фонемаларнинг у ёки бу сўз позициясидаги қўлланиш миқдорини турли стиль материаллари асосида қиёслаш ва бу орқали у ёки бу фонема даражасида сўз позицияси бўйича турли стилларни бир-бирларидан фарқлаш ёки бир хиллик ҳолатлари мезонини аниқлаш мумкин.

Лингвостатистик тадқиқот учун олинган тўртгала хил матндаги унли фонемаларнинг сўзнинг ҳар бир позицияси (боши, ўртаси, охири)даги фониз миқдорларини алоҳида-алоҳида қўшиб, умумий стиль сонига бўлсак, ҳозирги ўзбек тили матнларида унли фонемаларнинг сўз позициясидаги статистик тақсимланиши намоён бўлади. Булар асосида тегишли параметрлар тавсия этилди.

Ҳозирги ўзбек тилида барча унли фонемалар сўз боши ва сўз охиридагига нисбатан сўз ўртасида кўпроқ қўлланади. Бинобарин, ҳозирги ўзбек тили унлилар таркибидagi барча унлиларнинг ёшасига сўздаги позициясига кўра миқдорий муносабати қуйидагича – 10 : 74 : 16, яъни сўз бошида (6945 : 72012) – 10 %, сўз ўртасида (53186 : 72012) – 74%, сўз охирида (11881 : 72012) – 16%. Шунингдек, сўз боши билан сўз охиридагиларининг ўзаро миқдорий муносабати – 37(36,8904706) : 63.

Умуман, кузатишган ҳамма стилларда ҳам унлиларнинг қўлланиш частотасига кўра сўздаги ўрни бўйича статистик жойлашиш даражаси аниқланди. Унлилардан о, у, е /э/, ў лар сўз охирида 3-ўринда қўлланиши, ҳозирги ўзбек тилида ўша

фонемалар билан тугаган сўзларнинг камлиги билан изоҳланади. Булар, асосан, бошқа тиллардан ўзлашган рўбарў, обрў каби сўзлардир.

Ҳозирги ўзбек тили унли фонемалари сўзларнинг турли позициялардаги қўлланиш характеристикасига кўра, яъни фонематик даражадаги статистик мезонларга биноан турли стилларни ўзаро фарқлаш мумкин.

Ҳозирги ўзбек тилидаги Н, Л, Р, М, Ҳ - сонорларининг проза матнларидаги салмоғи 22%, поэзия матнларидаги салмоғи 24%, ижтимоий-гуманитар матнларидаги салмоғи 23%, аниқ фанлар матнларидаги салмоғи 24 %. Демак, проза матнлари бўйича 11 сонангнинг умумий матндаги салмоғи 64%, поэзия матнлари бўйича 64%, ижтимоий-гуманитар матнлари бўйича 65% ва аниқ фанлар матнлари бўйича эса 66 % дир.

Лаб ундошлари (Б, М, В, П, Ф)нинг проза матнларидаги салмоғи (F = 4012) - 16%, поэзия матнларидаги салмоғи (F=3946) - 16%, ижтимоий-гуманитар матнларидаги салмоғи (F=3990) - 16%, аниқ фанлар матнларидаги салмоғи (F=3865) - 15 %.

Ўзбек тилидаги жуфт ундошлар (Д-Т, Б-П, Г-К, З-С, В-Ф, Ҳ-Х, Ж-Ч, Ҳ-Ш)дан жаранглиларнинг жарангсизларга нисбати бўйича проза матнларида (барча фонемаларга нисбатан) - 26%: 25% ёки ўзаро 51% : 49%. Частоталари - 6385 : 6178; поэзия матнларида - 24%: 24% ёки 50% : 50%. Частоталари - 6153 : 6151; ижтимоий-гуманитар матнларида - 22% : 26% ёки 46% : 54%. Частоталари - 5576 : 6482; аниқ фанлар матнларида - 21% : 27% ёки 44% : 56%. Частоталари - 5358 : 6806.

Тилининг горизонтал ҳаракатига кўра орқа қатор унлилар (У, Ў, О) нинг олд қатор унлилар [И, Е(Э), А]га нисбатан миқдорий муносабати бўйича проза матнларида (барча фонемаларга нисбатан) - 11% : 31% ёки ўзаро 26% : 74%, поэзия матнларида - 12% : 28% ёки 30% : 70%, ижтимоий-гуманитар матнларида - 10% : 33% ёки 23% : 77%, аниқ фанлар матнларида - 9% : 33% ёки 21% : 79%.

Тилининг вертикал ҳаракатига кўра унлиларнинг (барча фонемаларга нисбатан) турли стилларда қўлланишининг % миқдори ишлаб чиқилди.

Турли матнларда ҳозирги ўзбек тили фонемаларининг таснифи бўйича “фонематик занжир”да фонемалар тақсимланишининг бошқа туркий тиллар билан келажакда нисбатан кўпроқ солиштирилиши мумкин бўлган баъзи бир томонларинигина келтирдик, холос.

Ҳозирги ўзбек тилида назарий жиҳатдан қўлланиши мумкин бўлган бирикмалар (икки фонемали)дан, яъни 1024 (“ Δ ” билан бирга $32^2 = 1024$)дан текширилган матнимиздаги 758 таси (“ Δ ” сиз $31^2 = 961$ дан 698 таси) учради.

Ўзбек тилидаги турли функционал стилларда ва умуман, тил бўйича фонемаларнинг симметриялилик муносабати диссертацияда келтирилган.

Лингвостатистик аспектда кузатишган матнларда учрамаган айрим бирикмалар танланма ҳажми (бизнинг материалимизда $N_1 = 42960$; $N_2 = 42603$; $N_3 = 43524$; $N_4 = 43389$; умумий $N=172476$ ни) янада давом эттириш (ҳар хил мавзуларга доир турлича матнлар ҳисобига тўлдирилиши) пировардида “янги” бирикмалар аниқланиши (“минус”лик ўринлар йўқолиши) мумкин. Ҳар ҳолда

ўшандай “янги” бирикмалар ўзбек тилида нисбатан камроқ учраши билан характерланади, холос .

Биз кузатган матнларда ҳозирги ўзбек тилида назарий жиҳатдан қўлланиши мумкин бўлган барча икки фонемали бирикмаларнинг 74%и учради. Бу эса икки фонемали бирикмалар бўйича баъзи бир статистик тавсифлар бериш учун қоникарли ҳисобланади.

Ҳозирги ўзбек тилидаги ҳар бир фонема барча фонеманинг неча % и билан жуфт ҳолда кўпроқ қўллана олиши лингвостатистик тадқиқ ва таҳлил этилиб, тегишли параметрлар тавсия этилади.

Ҳозирги ўзбек тилида ҳар иккала томондан ўзаро бирикмага (жуфтликка) эга бўлмаган фонемалар (яъни “а” билан “у” да “ау” ҳам, “ўа” типини ҳам учрамаган фонемалар)нинг саксон етти типини диссертацияда келтирилган.

Лингвостатистик аспектда кузатилган матнларда яна шундай икки фонемали бирикмалар - жуфтликлар маълум бўлдики, бундай бирикмалар таркибидаги фонемаларнинг ўрнини алмаштириш мумкин эмас (алмаштирилган ҳолдагиси учрамади: яъни “рц” учради, “цр” эса учрамади). Диссертацияда бундай икки фонемали бирикмалардан саксон тўққиз тури келтирилган.

Икки фонемали бирикмалар ҳаторида (нолга тенг “нулевая фонема”) – “пробел” ҳам “бир” фонема тарзида ҳисобга олинмаган бўлса, у ҳолда ΔС, СΔ, ΔV, VΔ типидагилар ҳам шартли равишда икки фонемали бирикмалар сифатида қаралади(3-жадвал).

3-жадвал

Фонемалар бирикмаларининг ўзига ҳослик даражаси

Бирик-малар	Назарий жиҳатдан қўлланиши мумкин бўлган бирикмаларнинг миқдори	Матнда учраган бирик-маларнинг миқдори	“В”нинг “Б”га нисбати (% ҳисобида)	Матнда учраган бирикмалар (“В”даги)нинг умумийга (Σ = 758га) нисбати (% ҳисобида)
А	Б	В	Г	Д
ΔΔ	-	-	-	-
VV	36	23	64 %	3 %
CC	625	388	62 %	51 %
VΔ	6	6	100 %	1 %
ΔV	6	6	100 %	1 %
CΔ	25	24	96 %	3 %
ΔC ¹	24	24	100 %	3 %
CV	150	143	95 %	19 %
VC	150	144	96 %	19 %
Σ	1024	758	74 %	100 %

¹ Ҳозирги ўзбек тилида “Н” фонемаси билан (Δ Н) сўз бошланмайди. Шунинг учун “ΔС” бирикмалар бўйича назарий жиҳатдан ҳам 25 ундوشин эмас, балки 24 ундوشин кўрсатдик.

Бунда сўз бошидаги ва сўз охиридаги ундошлар, шунингдек, сўз бошидаги ва сўз охиридаги унлилар қўлланишининг статистик параметрлари яққол кўринади. Ҳозирги ўзбек тилининг турли функционал стил(матн)лари бўйича фаол қўлланиладиган (частотаси 100 ва ундан ортиқ бўлган) икки фонемали бирикмалардан CC , CV , VC , ΔV , $V\Delta$, ΔC , $C\Delta$ типлари диссертацияда алоҳида-алоҳида қайд этилган.

Ҳозирги ўзбек тилидаги жуфт фонемаларнинг қанчалик характерлилик даражаси, шунингдек, у ёки бу бирикмалар(CC , VV , ...)нинг қанчалик чекланганлиги лингвостатистик жиҳатдан ишлаб чиқилди.

Фонологостатистик тадқиқот юзасидан лингвостатистик параметр сифатида қуйидагилар алоҳида қайд этилди:

Тилдаги олти унли ва олти сонор (а, о, и, у, э, ў, н, л, р, м, й, н)нинг умумий матндаги салмоғи 68 % лигидан англашиладики, ҳозирги ўзбек тилини фонематик таркиби жиҳатидан тўла равишда сонангли деб, тавсифлаш мумкин.

Ҳозир ўзбек тилининг “фонематик занжири”да энг кўп қўлланиладиган фонема (товуш) типи – CV (ундош + унли).

Энг кўп ва кам қўлланиладиган икки фонемали бирикмалар типини қуйидагича (Икки сўз орасидаги бўшлик, яъни пробел билан ҳисобга олинганда):

$$CV > VV, CV > CC, VC > VV, VC > CC, VV < \Delta V, \\ VV < V\Delta, CC < \Delta C, CC < C\Delta.$$

Икки фонемали бирикмалар, аксарият, сўз асосида (ёки ўзагида), шунингдек, ўзак + аффикс, яъни сандхи ҳолатда учрайди.

Ўзбек тилида геминатлар, асосан, сўз ўртасида қўлланади. Турли типдаги икки фонемали бирикмалар геминатларга нисбатан кўпроқ қўлланиши билан характерланади.

Турли типдаги икки ундош фонемали бирикмалар сўзнинг ҳамма ўрин(позиция)ларида қўллана олади. Хусусан, сўз бошида 10–15 %, сўз ўртасида 75–85 % ва сўз охирида 22–30 % ни эгаллайди.

Ҳозирги ўзбек тилида энг кўп қўлланиладиган икки ундош фонемали бирикмалардан $нд$, $рд$, $лд$, $лр$, $рг$, $ст$, $рн$, $тл$, $ил$, $вл$, $мт,гл$, $кт$, $кт$ ва бошқалар қайд этилди.

Диссертациянинг “Бўғиннинг статистик структураси” номли тўртинчи бобида бўғин типлари ва уларнинг статистик структураси, бўғиннинг фонематик тартиби, сўзнинг ўртача узунлиги каби масалалар тадқиқ этилди.

Г.П.Торсеевнинг фикрича, сўзнинг товуш структурасида субструктуранинг тўрт кўриниши (фонемали, фонемаларнинг бирикиш структураси, силлабик ва акцент - ритмик структура) ўзаро таъсир қилади ва бирлашади. Улардан силлабик структурада бўғин ёки сўз ясайдиган бўғиннинг аниқ кетма-кетлиги тасаввур

қилинади. Силлабик структура бўғиннинг миқдорини, унинг типни ва таркибини, кетма-кет тақсимланиши ва бўғин бўлинишини белгилайди¹.

Ўзбек тили сўзшаклларида бўғинлар сони 1-10 гача ёки ундан ҳам ортик бўлиши мумкин. Айниқса, оғзаки ёки ёзма нутқда неча бўғинли сўзлар кўп қўлланишини аниқлаш қимматлидир.

Ўзбек тилининг оригинал бадий-прозаик стили бўйича олиб борилган кузатишлар, бундай стилда сўзнинг ўртача узунлиги уч бўғин сонига яқинлашишини кўрсатди.

Инглиз тилида 1-8 бўғинли сўзлар мавжуд бўлиб, булардан 1-4 бўғинли сўзлар энг кўп қўлланиши билан ажралиб туради.

Жоунзнинг талаффуз луғатида, — деб кўрсатади Г.П. Торсуев, — етти бўғинли сўзлар ўнлаб ҳисоблангани ҳолда, саккиз бўғинли сўз эса биттадир².

Умуман, муайян тилдаги бўғин сонлари сўзнинг товуш структураси хусусиятларини тавсифлашда, шунингдек, турли тилларни киёсий-типологик ўрганишда катта аҳамият касб этади.

Ўзбек тилида бўғин структурасини лингвостатистик таҳлил қилиш учун ҳозирги ўзбек тилидаги мавжуд фонемалар системасига (фонологик транскрипцияга) асосландик.

Умуман, тил дистрибуцияларини ўрганиш учун энг қулай объект фонологик бўғиндир.

Бўғинни лингвостатистик тадқиқ ва таҳлил этишда бўғиннинг статистик структураси (қўлланиш частотаси) билан бўғин структурасининг статистикасини чалкаштирмаслик керак. Иккаласи ўз ўрнида таҳлил этилса, аини мудаодир. Тадқиқ жараёнида буларга алоҳида эътибор берилди.

Буларнинг сўздаги ўрни (сўз боши, сўз ўртаси ва сўз охири) бўйича қўлланиш частотаси ҳам махсус тадқиқ ва таҳлил қилинди.

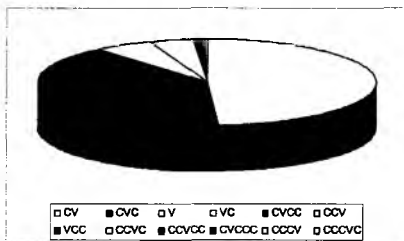
Ўзбек тилининг оригинал проза матнларида бўғин типлари (бўғиннинг фонематик таркиби)нинг 12 хил бўғин типни (CV,CVC,V,VC, CVCC, CCV, VCC, CCVC, CCVCC, CVCCC, CCCV, CCCVC)дан ташқари бўғиннинг VCCC (Омск), VCCCC (Эрнст), CCVCCC (Днестр), шунингдек, CCCCCV, CCCCCV, CCCCVC, CCCVCC, CVCCCC, VCCCC каби типлари ўта кам қўлланиши билан характерланади(4-жадвал).

¹ Торсуев Г.П. Проблемы теоретической фонетики и фонологии. - Л.: Наука, 1969. - С. 6.

² Қарағ: Г.П.Торсуев, кўрсатилган асар. - С. 70; Daniel Jones. Everyman's English Pronouncing Dictionary. - London-New York, 1957.

Бўгин тиллари ва уларнинг қўлланиш частотаси

Бўгин тиллари	Фониз хисобида
CV	48,983
CVC	39,967
V	5,610
VC	4,097
CVCC	0,830
CCV	0,177
VCC	0,173
CCVC	0,137
CCVCC	0,013
CVCCC	0,007
CCCV	0,003
CCCVCC	0,003
Σ	100,000



Ўзбек тилининг оригинал бадиий прозаик матнларида бўгин тилларининг дастлабки тўрт тиби (CV, CVC, V, VC) муҳим ҳисобланган, яъни улар умумий матн (бўгин) ҳажмининг 99 % (аниқроғи - 98,657) ни ташкил этган.

Демак, миқдорий далиллардан келиб чиқсак, у ҳолда бўгиннинг (сўздаги ўрни ҳисобга олинмаган ҳолдаги) CV, CVC, V, VC тилларини ўзбек тилининг оригинал бадиий-прозаик стили учун характерли, деб баҳоланиши мумкин. Шунингдек, сўзлар структурасини ташкил этишда бўгиннинг CV ва CVC тиби фаол қўлланади.

Аксарият тилларда CV бўгиннинг кенг тарқалган тиби бўлганлиги учун бўгин конструкциясининг элементар модели ҳисобланади.

Ҳозирда туркий тиллар, хусусан, ўзбек тилида, рус тили ва бошқа чет тилларнинг самарали таъсири натижасида бўгин тилларининг сони кўпайиб бормоқда. Н.А.Баскаков кўпчиликл туркий тилларда 6 асосий тип (2 очик бўгин: V, CV; 4 ёпик бўгин: VC, CVC, VCC, CVCC) мавжудлигини кўрсатади¹. Ўзбек тилининг оригинал бадиий-прозаик стилида бўгинларнинг 90 % ундош фонема билан бошланади, қолган 10 % эса унли фонема билан бошланади. Ундош фонема билан бошланадиган бўгин тиллари (C ...)нинг унли фонема билан бошланадиган бўгин тиллари (V ...)га нисбатан миқдорий муносабати - C : V ≈ 9.

Ўзбек тилининг оригинал бадиий прозаик матнларида бўгиннинг V, VC, CV тиллари сўз бошида (сўз охирида ёки сўз ўртасида эса бўгиннинг CVC тиби) энг

¹ Қараг: Баскаков Н.А. Структура слога в тюркских языках / Сб. "Исследования по сравнительной грамматике тюркских языков". Ч.1 (Фонетика). - М., 1955. - С.333-334; Баскаков Н.А. Историко-типологическая фонология тюркских языков. - М.: Наука, 1988; Шунингдек, бўгин тиллари масаласи хақида яна қараг: Шеворопкин.В.В. Звуковые цепи в языках мира. - М., 1969. -С.32-62.

кўп қўлланиши билан характерланади. Бўгиннинг сўздаги ўрни бўйича статистик тақсимланиши, шунингдек, бўгиннинг абсолют ёки нисбий частотасининг аниқланиши бўгиндан сўз ҳосил қилиш жараёнини лингвостатистик тадқиқ этишда В. Фукс формуласини татбиқ этишга йўл очади. Немис математиги В. Фукс формуласида бўгинлардан сўз ҳосил бўлиш жараёнини математик тасвирлаш мумкинлиги ифодаланади. В. Фукс ўтган асрнинг 50-йиллари охирида лингвистик элементларни (кўпгина тилларда элемент сифатида 1,2,3 ва ундан ортиқ бўгиндан ташкил топган сўзлардан фойдаланилган) уларнинг компонентларидан ясалишининг математик назариясини яратган¹.

Бўгинлар тарқибдаги фонемаларнинг миқдори қуйидагича: икки фонемали (53,080%), уч фонемали (40,317%), бир фонемали (5,610%), тўрт фонемали (0,970%), беш фонемали (0,023%). Англашадикки, ўзбек тилида икки ва уч фонемали бўгин типни ҳам характерлидир. Ҳозирги ўзбек тилидаги ҳар бир фонема билан бошланадиган “бўгин хиллари” нинг миқдори аниқланди.

Ўзбек тилининг оригинал бадиий-прозаик стилида сўзнинг ўртача узунлиги - 2,572 бўгинга тенг. Бундай бўгин сонини қуйидаги формула асосида ҳисоблаш мумкин:

$$P(i) = P_i = \frac{Z_i}{k},$$

бундан, $i \left(1 \leq i \leq k; \sum_{i=1}^k P_i = 1 \right)$ - сўздаги бўгин сони;

k - матндардаги сўзлар сони. Агар i - сони мураккаб бўлса, яъни “ i ” бўгиндан ташкил топган бўлса, у ҳолда “ Z_i ” билан белгиланади.

Демак, матндаги P_i нинг қиймати, улуши - $P_i = \frac{Z_i}{k} = \frac{30000}{11664} = 2,572$.

Сўздаги бўгиннинг ўртача миқдори ёки тақсимланишининг ўртача қиймати -

$$\bar{i} = \sum_{i=1}^k i \cdot P_i.$$

С.Г.Чебановнинг фикрича, турли тилларда бўгинларнинг сонига қараб сўзларнинг тақсимланиши баъзи бир умумий қонуниятларга бўйсунади².

Тилдаги очиқ бўгинларнинг ёпиқ бўгинларга нисбатан миқдорий муносабати - 54,773 : 45,227.

¹ Қарағ: Fuchs W. Mathematical Theory of word – Formation, Physics Institute of the University of Technology (Aachen, 1955); Яна қарағ: Мерките Р. Некоторые статистические характеристики образования слов из слогов и слогов из букв для литовского языка // “Литовский математический сборник”. II № 1. – Вильнюс, 1962. – С.91 – 105.

² Қарағ: Чебанов С.Г. О подчинении речевых укладов “индоевропейской” группы закону Пуассона // Доклады АН СССР. Новая серия. Т.55. - № 2. - 1947. – С.103 - 108.

Унли (V) ва ундош (C) ларининг ўзаро миқдорий муносабати - консонантли коэффициент¹ - $N_c : N_v \approx 1,37$.

Ҳозирги ўзбек тили сўз таркибидаги биринчи бўғинни аниқ билиш орқали иккинчи бўғин ҳақида маълум бир ахборотга эга бўлиш мумкинлиги муаммоси ҳали ўзининг охири ечимигача етганича йўқ.

Шуни таъкидлаш керакки, ўзбек тилининг оригинал бадий-прозаик стилидаги бўғин статистик структурасининг бошқа стиль материаллари билан қиёсланиши орқали ҳар бир стильга оид (бўғин даражада ҳам) статистик мезонининг аниқланиши, шубҳасиздир. Бинобарин, лингвостатистик маълумотлар бўғин даражасида турли стилларнигина эмас, балки турли ёзувчи ва шоирлар услубларини бир-бирдан чегаралаш имконини беради.

Демак, ҳозирги ўзбек тилининг оригинал бадий-прозаик стилида икки ва уч фонемали бўғин типини характерлидир. Уларнинг умумий матндаги салмоғи 93,397 %.

Бўғин типларини ҳосил қилишда Р, Л, Н, Ё ундошлари фаоллиги билан ажралиб туради.

Олинган танланма ҳажми доирасида 1294 бўғин хили аниқланиб, булардан 58 хили умумий матн ҳажмининг ярмидан кўпроғи(56,386667 %)ини ташкил этади.

Ҳозирги ўзбек тилида бўғиннинг CCV(сме-на), CVCC(қирқ), CCCV(стра-те-ги-я), CCVC(врач), VCCC(Омск), VCCCC(Эрнст), CVCCC(пункт), CCVCC(спорт), CCCVC(скрип-ка), CCVCCC(Днестр) тиллари рус тили ва у орқали ўзлашган сўзлар таъсири маҳсулидир².

Диссертациянинг бешинчи боби “Лексик статистика”³га бағишланган бўлиб, унда сўз ва сўзшаклларининг дастлаб морфематик структураси, морфемаларнинг ўринлашуви, морфемаларнинг фонологик структураси борасида қилинган ва келгусида бажарилиши лозим бўлган муаммолар ўртага ташланган. Шунингдек, болалар адабиёти тили лексикаси, лексик статистиканинг ўзига хос хусусиятлари, олинган танланмадаги янги сўзларнинг қўлланиш частотаси, юқори частотали сўзлар, сўз туркумидаги фаол сўзлар, сўзларнинг этимологик жиҳатдан тақсимланиши, сўз(сўзшакл)лар ва уларнинг статистик структураси, лексикографиядаги лингвостатистик тадқиқотлар, Ҳ.Ҳ.Ниёзий “Девон”и тили лексикаси (сўз шакллари) нинг статистик структураси тадқиқ ва таҳлил этилди.

Сўзнинг морфологик структураси нуктаи назаридан бош+ла+в+чи+лар+имиз+нинг сўзи етти морфемадан иборат, деб қаралган.

¹ Жаҳон тилларининг “товуш занжири”да унлилар миқдорининг ундошлар миқдорига бўлган муносабати ($N_v : N_c$) - 2 ва 0,5 орасида тебранади. Қарағ: Шеворошкин В.В. О структуре звуковых цепей /В. кн. “Проблемы структурной лингвистики”. - М., 1963. - С.164 - 181.

² Назарий жиҳатдан қўлланиши мумкин бўлган CCCCV, CCCCCV, CCCVC, CCVCC, CVCCCC, VCCCC бўғин типлари ҳозирча бизнинг кузатган матнларимизда учрамади. Бу тиллар учраган тақдирда ҳам, пассив ҳисобланади.

³ Сўзшакллари статистикаси (сўзшакллари қўлланиш частотаси)ни шартли равишда лексик статистика деб умумлаштирилиб, унинг ичида ҳозирча сўзшаклларининг морфем структураси ҳам келтирилди.

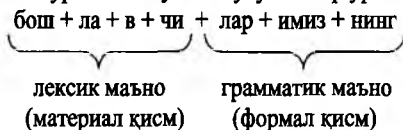
Аслида сўзни бош+ла+в+чи+лар+имиз+нинг тишида таҳлил қилиш, сўзнинг морфемалар иштирокидаги структураси сўз ясалиши структурасига ҳам, морфологик структурасига ҳам тегишли бўлмай, балки сўзнинг морфем (морфемали) структурасига, қолаверса, тилшуносликнинг алоҳида мустақил соҳаси ҳисобланган “Морфемика”га тааллуқлидир¹.

Эркин ўзаклар ўзлари аффиксларсиз бошқа морфемалар билан бирлашмаган ҳолда ҳам мустақил қўлланилаверади. М.: “вагон”, “вагонлаб” ва ҳ.к.

Шартли (шартланган-боғли) ўзаклар эса тилда фақат аффикслар билан birlikда қўлланади. Улар, асосан, рус ва рус тили орқали ўзлашган сўзлар таркибида кўпроқ учраши билан характерланади. Масалан: тирмаламоқ, тирмашмоқ феълларида тирма учрамайди. Қуйидаги сўзлар ҳам аффикссиз қўлланмайди: тур+изм, тур+ист, ирриг+атор, машин+ист, музик+ант, ате+изм, ате+ист, дег+ча, шол+ча. Бу кабилар шартли (боғли) ўзаклардир.

Морфем луғат мавжуд, аммо лингвостатистик аспектда сўзларнинг турли функционал стилларида қўлланиш частотаси нуқтаи назаридан морфем статистика хануз яратилгани йўқ.

Умуман, морфемикада, сўзнинг морфемалар таркиби, морфемалар тури ва ҳар бир морфема турига хос бўлган хусусиятлар ўрганилади. Масалан:



Айни мана шу иккала қисмининг, яъни сўзнинг тилда мавжуд бўлган материал қисми билан, нутқда мавжуд бўлган формал қисмининг умумий бир системада ўрганилиши тилшуносликнинг алоҳида бир мустақил бўлими (соҳаси) ҳисобланмиш морфемикага тегишлидир.

Морфемика ҳам сўз ясалишига, ҳам морфологияга “қириш”дан иборат.

Тил системасида сўзнинг энг кичик (бўлинмас) маъноли қисмлари ҳисобланган морфемалар ўзларининг асосий хусусиятларига кўра икки тип (тур)га бўлинади:

1. Асосий (ўзак) морфема;
2. Ёрдамчи (эргаш, аффиксал) морфема (аффикс)².

Асос (ўзак) морфема - шу сўзнинг асосида ётган лексик маъно ифодалай оладиган қисмидир; сўз яшаш учун асос бўладиган қисмдир.

Демак, ўзак морфлар йиғиндисининг умумлашгани ўзак морфемадир³.

¹ Грамматика современного русского литературного языка. - М., 1970. - С.30; Ўзбек тили грамматикаси (коллектив), I том (морфология). - Т.: Фан. 1975. - Б.11.

² Современный русский язык. Морфология. Под редакцией акад. В.В.Виноградова. - М., 1952. -С.34; О.П.Супик. О морфологическом составе слова в ... языках. - Изв. АН СССР, ОЛ. Я. 1958. - XVII том, 4-вып. - С.331 - 336.

³ Ўзак шу сўзнинг асосий лексик маъносини билдирган, морфологик жиҳатдан бўлинмайдиган элемент, сўзнинг адриси: ўзаксиз сўз йўқ (проф. А.Фулумов). Қараг: Ўзбек тили грамматикаси (коллектив)..., - Б.69.

Аффиксал морфема - сўзнинг лексик маъно ифодалай олмайдиган, сўзнинг лексик ёки грамматик маъносининг шаклланишида иштирок этадиган қисмидир.

Аффиксал морфларнинг умумлашгани эса аффиксал морфема ҳисобланади. “Бошловчиларимизнинг” да: бош – асос(ўзак) морфема; -ла, -в, -чи, -лар, -имиз, -нинг - аффиксал морфема.

Ҳинд-европа тилларида сўз таркиби таҳлили “асос”дан бошланса, туркий тилларда, хусусан, ўзбек тилида ўзакдан бошланади. Баъзан ўзакка аффикс қўшилганда ўзакнинг фонетик структурасида турли хил ўзгаришлар рўй бериши мумкин. Бундай ҳодиса тилшуносликда морфонологик (ёки фонеморфологик) ҳодиса номи билан юритилади. Морфонологияга тегишли ҳодисаларнинг ҳаммаси морфемика бўлимининг ичида тадқиқ этилади.

Ҳозирча тилшуносликда умумтомондан қабул қилинган ва таърифлаб келинаётган “Морфема – сўзнинг ўзига хос маънога эга бўлган энг кичик бўлинмас маъноли қисм”лиги юзасидан фикр-мулоҳазалар етакчилик қилмоқда.

Сўз структураси морфем таҳлил қилинганда сўзларнинг асосий ва ёрдамчи морфемаларга бўлиниши ва уларнинг турли аспектдаги таҳлили тушунилади (масалан, ўзак ёки аффиксларнинг фонетик – фонологик структураси каби).

Сўзшаклланишнинг таркибида барча морфемаларнинг умумий статистик структураси ҳанузгача махсус тадқиқ этилгани йўқ.

Туркийшуносликда ва ўзбек тилшунослигида ҳам морфемика бўйича А.Н.Кононов, Н.А.Баскаков, Э.В.Севортян, А.Фуломов, А.Ҳожиёв, С.Усмонов ва бошқаларнинг илмий тадқиқотлари мавжуд. Ё.Тожиёвнинг “Ўзбек тили морфемикаси” (1992), “Ўзак морфема ва сўз муносабати”, “Ўзак морфемаларнинг фонетик тузилиши”, “Аффиксал морфема”, “Аффиксал морфемаларнинг фонетик хусусиятлари”, “Содда ва қўшма аффикс морфемалар” сингари изланишларида, Т.Мирзакуловнинг “Грамматика ўқитишнинг лингвистик асослари” (1994) номли рисолида, А.Абдуазизов, А.Нурмоновларнинг “Морфонология”га оид китоблари (Т., 2002)да “Морфемика ва унинг асосий масалалари”, “Морфемалар таснифи”, “Ёрдамчи морфемалар тавсифи”, “Морфонология ва морфосинтаксиснинг баъзи масалалари” қисмларида морфемиканинг лингводидактик ва тарбиявий асослари, асосий тушунчалари ва бирликлари, морфемаларнинг маъно ҳамда вазифаси билан боғлиқ ҳолда таснифи масалалари, морфемалардаги товуш ўзгаришлар ва уларнинг бириктиш хусусиятлари ҳақида илмий-назарий ва амалий маълумотлар келтирилганлигига қарамастан, лингвостатистик метод асосида “Морфемаларнинг ўринлашуви”, “Морфем структура”, “Морфемаларнинг фонологик структураси” кабиларнинг лингвостатистик тадқиқи долзарблиги билан ажралиб туради. Айниқса, А. Ҳожиёвнинг “Ўзбек тили ва адабиёти” журнали (1998 йил 2-сонидан эътиборан) чоп этилган 10 дан зиёд туркум мақолалари долзарблиги билан кўпчиликнинг диққатини қозонди. Маълумки, лексема - лексик birlik, сўз – лексик-грамматик birlik. Сўз (лексема каби) тил қурилиши birlikларининг иккинчи турига оид birlikлар каби вазифа бажаради, - деб ёзади А.Ҳожиёв, - тил

қурилиши – морфологик қурилиши сатҳига оид бирликнинг (сўзшаклининг) ясалишида асос (ясалиш асоси) бўлиб хизмат этади.

Тил қурилиши асосий бирликларининг тўғри ва тўлиқ ёритилишида луғавий ва грамматик морфеманинг моҳиятини илмий-назарий нуқтан назаридан тўғри белгилаш, морфеманинг қандай бирлик ҳисобланиши А. Ҳожиёвнинг ўша туркум мақолаларида ўз ифодасини топган.

Сўзшакллари частотали луғати асосида сўз ёки сўзшаклларини морфемаларга ажратишда ҳозиргача амалда қўллаб келинаётган тартибдан фарқли ўлароқ чап томондан (асос томондан) ўнг томонга қараб эмас, балки ўнг томондан — сўзшаклидаги охириги морфемадан чап томонга қараб бўлиниши, яъни морфемаларга ажратилиши мақсадга мувофиқдир. Бу эса морфемага берилган умумий таърифга тўла мос келади.

Лексикостатистиканинг марказий муаммоси жонли тилнинг статистик қонуниятлари ва матннинг статистик структурасини аниқлашдир. Матннинг статистик структураси дейилганда, шартли равишда маълум матн (яъни сўзлик ҳажми)даги турли сўзлар миқдори билан матнда унинг қайтарилиш частотаси орасидаги муносабат тушунилади, яъни матннинг барча сўзлари янги бўлиши мумкин: $N = L$, ёки сўзнинг қисми такрорланади $N > L$. Ҳар минг сўзда қанча янги сўз борлиги шу тариқа тадқиқ ва таҳлил қилинди. Ҳар минг танланмадаги янги сўзларни ўзидан аввалгиси билан қиёслаш орқали маълум бир умумий танланмада янги сўзлар миқдорини аниқлаш мумкин. Бинобарин, танланма ҳажми қўлайиб бориши ($N \rightarrow \infty$) билан янги сўзлар миқдорининг камайиш ($P \rightarrow 0$) жараёнини ўрганиш мумкин.

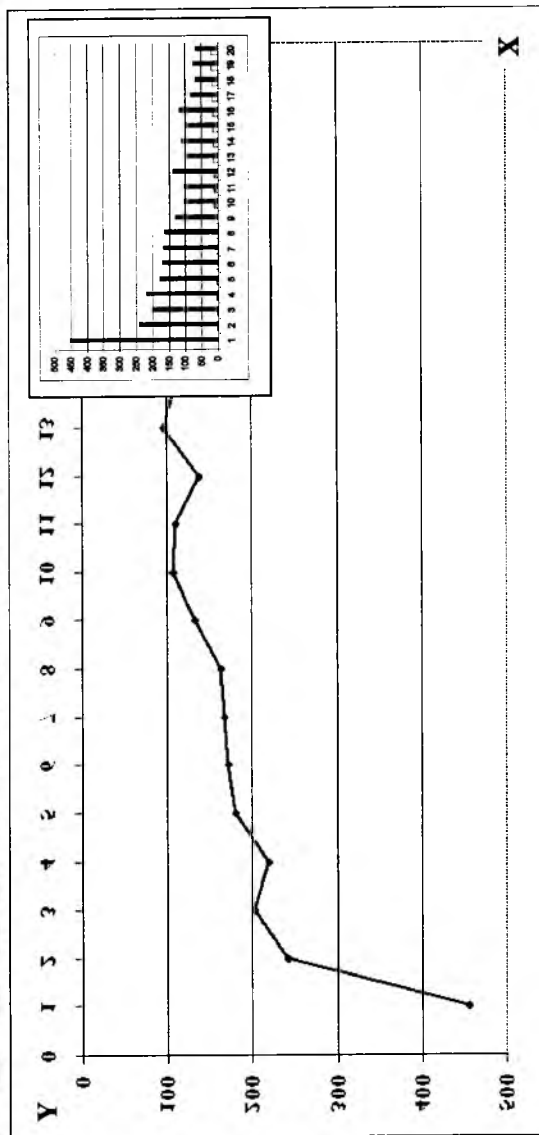
Лекин бу ҳолат бирор тилдаги жами янги сўзлар миқдорини аниқ белгилаш деган гап эмас. Чунки Y - ордината чизиги томонидан келаётган синиқ чизиқ, маълум вақтда - қониқарли танланма ҳажмида X -абсцисса чизиги билан учрашиши, маълум вақт ўша чизиқда давом этиши ва ундан яна юқорига кўтарилиши мумкин (5-жадвал ва 1-диаграмма).

Англашиладики, биринчи мингликда янги сўзлар миқдори 455 бўлса, ўнинчисида - 107, ўн бешинчисида - 98, йигирманчисида эса - 67, яъни матн миқдорининг ошиб бориши билан сўзлик ҳажмининг ўсиш суръати камайиб борапти. Бинобарин, матн миқдори билан сўзлик ҳажми орасида боғланиш мавжуддир. Бироқ бу боғланишни кўпроқ кичик ҳажмли танланмалардагина яққол кўриш мумкин.

Сўзшакллар қўлланиш частотасининг дастлабки лингвостатистик тадқиқи юзасидан ҳар минг танланмада янги сўзшакллар миқдори ҳамда ҳар минг сўз(сўзшакл)даги янги сўзларнинг ўртача частотаси аниқланди. Буларга суянган ҳолда статистик таҳлил амалга оширилди. Авторефератнинг 14-бетида келтирилган эҳтимол хато — $L = \frac{t}{\sqrt{n}}$ дан фойдаланилди.

Хар минг сўздаги янги сўзлар миқдори

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Жами
Y	455	241	202	219	180	171	167	163	132	107	110	138	95	112	98	118	84	69	76	67	3004



1- диаграмма

Бунда t - константа (95 %лик ишновч учун 20 таъланмада $t=2,09$), $\sigma = \sqrt{\frac{\sum a_i^2}{n}}$ - ўртача

квадратик фарқ, баъзан σ ўрнида S дан $\left(S = \sqrt{\frac{\sum a_i^2}{n-1}} \right)$ фойдаланиш яхшироқдир. Демак,

$$L = \frac{t \cdot \sqrt{\frac{\sum a_i^2}{n-1}}}{\sqrt{n}} = \frac{2,09 \cdot \sqrt{\frac{6715,803}{19}}}{\sqrt{20}} = 9.$$

Қўринадик, шундай ҳажмдаги таъланмалар устида битта тажриба эмас, балки 100 та тажриба ўтказилса, шундан 95 тасида ўрта частота (445, 463) ораликда бўлиб, қолган беш тажриба мазкур ораликдан чиқиши мумкин.

Ўқоридаги ҳолат абсолют хато номи билан ҳам юритилади. Абсолют хато билан бирга нисбий хатоми аниқлаш керак. Абсолют ва нисбий хатоларни аниқлашдан мақсад қайси ҳолатда хато катта ва қайси ҳолатда кичик эканлигини билишдир. Нисбий хато $\delta = \frac{t \cdot \sigma}{\bar{x} \sqrt{n}}$ формула асосида аниқланди, яъни

$$\delta = \frac{2,09 \cdot 18,32}{454,25 \cdot 4,472} = 0,02 = 2\%.$$

Ўртача частота олиниб, статистик ишлаб чиқилгандан сўнг лингвостатистик қонун тўғрисида фикр юритишдан олдин ўртача квадратик фарқни ҳам аниқлаб олиш керак бўлади. Математик статистикада ўрта квадратик фарқ $\sigma = \sqrt{\frac{\sum a_i^2}{n}}$ формула ёрдамида аниқланади, бунда $a_i = x_i - \bar{x}$. Тадқиқотда ҳар минг сўзда қанча янги сўз борлигининг ўртача частотасини аниқлаш мақсадида тенг ҳажмли 20 та таъланма олинган эди. Шунинг учун биз ҳам юқоридаги формула асосида иш олиб боришни ўринли деб ҳисобладик.

Ўртача квадратик фарқнинг ўртача частотага бўлган фоизли муносабати (вариация коэффициентини) ўша фарқнинг фоизли ифодасидир. Вариация коэффициентини авторефератнинг 14-бетига келтирилган формула асосида аниқланди: $V = \frac{100 \cdot \delta}{\bar{X}}$ Бизнинг материалимизда $\sigma = 18,32$, $\bar{X} = 454,25$. Демак,

$$V = \frac{100 \cdot 18,32}{454,25} = 4\%.$$

Қўринадик, ўртача квадратик фарқ ўртача частотанинг 4%ини ташкил этар экан.

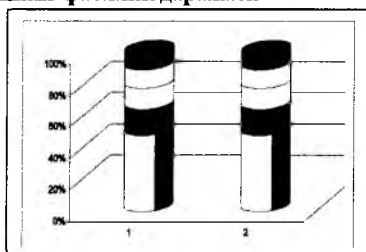
Умуман, вариация коэффициентини орқали ўзининг характери, катта-кичиклиги ёки миқдори билан турлича бўлган тил ҳодисалари частотасининг тебраниш даражасини қиёслаш мумкин. Вариация коэффициентини қанчалик кичик бўлса, унинг (текширилаётган ҳодисанинг) тебраниши шунчалик кичик бўлади.

Мазкур услуб тили лексикаси бўйича бир марта қўлланган сўзларнинг миқдор жиҳатидан кўплиги лингвостатистик тадқиқ орқали белгиланди. Бу эса ҳар қандай частотали луғатда (текширилаётган матн ҳажмига боғлиқ бўлмаган ҳолда) бир марта учраган сўзлар икки ва ундан ортиқ учраган сўзларга қараганда доимо кўпроқ қўлланишни тақозо этди (6-жадвал). Турли тилларнинг частотали луғатларида тез-тез учраб турадиган миқдор жиҳатидан у даражада кўп бўлмаган сўзлар гуруҳи (масалан, 50–60 таси) умумий матн ҳажмининг катта қисмига (50 % атрофига) тўғри келди. Ҳақиқатан ҳам, юқорида кўрсатилган рўйхатдаги сўзларнинг қўлланиш частотаси 52 ва ундан юқори бўлган 59 та сўз ($F=7032$) умумий матн ҳажмининг 54 % ини ташкил этди. Демак, ҳамма тилларда ҳам умумий матн ҳажмининг ярмидан кўпроғини ўша дастлабки 50-60 сўз ташкил этади. Бу ҳолатни частотали луғатлар асосида аниқ текшириб кўриш мумкин.

6-жадвал

Янги сўзшакллари миқдорининг фаоллик даражаси

Янги сўзларнинг неча марта учраши	Фоиз ҳисобида
1	47,90
2	17,24
3 - 4	13,32
5 - 9	11,42
10 ва ундан ортиқ	10,12
Жами	100,00



Сўзларнинг камайиш тартибида қўлланиш частотаси ва манбаларда учраши бўйича берилиши уларнинг қўлланиш частотаси нуқтаи назаридан ўзаро қиёслаш ва мазкур жанр(стил)га доир турли миқдорий қонуниятларни аниқлаш имконини беради. Шунингдек, тил нормаларининг объектив мезонини белгилаш учун ҳам ёрдам беради. Бундай тартиб асосида тузилган рўйхатда сўзлар ҳажмининг рақами билан унинг частотаси орасидаги боғланишни - Ципф қонуни ($P_z=k_z^{-1}$)ни ўрганиш мумкин¹. Бирок бу боғланиш маълум частота интервалидаги энг юқори частотали сўзлар доирасидагина мақсадга мувофиқдир.

Частотаси бўйича энг кўп ва фаол қўлланган сўзлар қуйидагилар:

бўлмоқ, 1, демоқ, олмоқ, келмоқ, қолмоқ, қилмоқ, кетмоқ, турмоқ, 1, чиқмоқ, кўймоқ, бормоқ, бермоқ, кўрмоқ, қарамоқ, тушмоқ, айтмоқ каби феъллар;

бола, ойи, кўз, қўл, гап, 1, иш, 1, кун, уй, буви, ака, ер, бош, мактаб каби отлар;

у, мен, бу, ўз, шу, нима, сен, ҳамма, биз, ким каби олмошлар;

кейин, энди, яна, жуда, шундай, ҳозир, кўп каби равишлар;

ҳам, 1, деб (2), лекин, ва, 1, деган (2), эса каби боғловчилар;

эди, йўқ, экан, бор, 1, ҳа, керак, эмас, майли каби модал сўзларлар.

¹ Фрумкина Р.М. К вопросу о так называемом "закоме Ципфа" // Вопросы языкознания. – М., 1961. - N 2. – С.117–122.

Ўзбек тили болалар адабиёти материаллари бўйича аниқланган юқори частотали сўзларнинг сўз туркумлари бўйича частотали тавсифи ишлаб чиқилди ва параметрлари аниқланди.

Болалар адабиёти матнида сўзларнинг қўлланиш коэффициенти ва “фойдаланиш” коэффициенти нуктаи назаридан боғловчилар, олмошлар, феъллар ва бошқалар характерлидир.

Танланма ҳажмини белгилашда математик статистиканинг тўғрилиқни баҳолаш методидан фойдаланиш ҳам илмий, ҳам амалий жиҳатдан эътиборга моликдир. Бу эса текширилаётган тил ҳодисаси частотасининг тўғрилиқ даражасини аниқлаш учун танланма ҳажми қанча бўлиши кераклигини белгилаб беради. Бунинг учун қуйидаги формуладан фойдаланилди:

$$\sigma = \frac{Z_p}{\sqrt{N \cdot P}}, \text{ бунда } \sigma - \text{нисбий хато; } Z_p - \text{константа (} Z_p=2\text{); } N - \text{танланма}$$

ҳажми; P – частота.

Камайиш тартибида тузилган частотали рўйхатдаги энг кам частотали ($F=10$) сўзларнинг частотасини ишончли натижа даражасига кўтариш учун, аввало, ўша частотали сўзнинг нисбий хатоси формула асосида аниқлаб олинди.

$$\text{Бизнинг материалимизда } N=13089; P = \frac{10}{13089}.$$

$$\text{Демак, } \delta = \frac{2}{\sqrt{N \cdot P}} = \frac{2}{\sqrt{10}} = 62\%.$$

Бундан кўринадики, нисбий хато частота кам бўлгани учун жуда катта. Нисбий хатони камайтириш учун танланма ҳажмини катта олиш керак. Ишончли натижанинг нисбий хатоси 20 % дан ошмаслиги шарт қилиб қўйилади. Бу - текширилаётган соҳа лексикасининг 80 % мазкур материал натижаларидан иборат эканлигини кўрсатади. Умуман, частотали лўғатлар лўғат бойлигининг 80 % ини қамраб олса, у қоникарли деб баҳоланади. Нисбий хатонинг частотага нисбатан, яъни δ нинг P га нисбатан тебраниш даражаси 0,0005

$$\left(\frac{P \cdot \delta}{100} \approx 0,0005 \right).$$

Бундан қуйи чегара $0,0008 - 0,0005 = 0,0003$; юқори чегара $0,0008 + 0,0005 = 0,0013$. Танланма ҳажми (N) ни белгилашда P ўрнида қуйи чегара (0,0003)дан фойдаланилади:

$$0,2 = \frac{2}{\sqrt{N \cdot 0,0003}}, \text{ бундан } N = \frac{4}{0,04 \cdot 0,0003} = 300000$$

Демак, частотаси 10 дан иборат бўлган сўзларнинг частотасини 60-70 лик даражасига кўтариш учун танланма ҳажмини 300000 дан кам олмаслик керак.

Агар $\delta = 10\%$ ликка эришмоқчи бўлсак, у ҳолда танланма ҳажми 1000000 ($0,1 = \frac{2}{\sqrt{N \cdot 0,0003}}; N = \frac{4}{0,01 \cdot 0,0003} = 1000000$)дан кам бўлмаслиги керак.

Нисбий хатони қанчалик кам олиш учун, яъни натижаларга тўла ишонч ҳосил қилиш учун танланма ҳажми катта ва хилма-хил бўлиши керак.

“Синчалак” повести тили сўзшаклларининг частотали лугатида сўзлар алифбо тартибда жойлаштирилишидан ташқари уларнинг матндаги қўлланиш миқдори - частотаси ҳам келтирилди.

Лугатда 100000 сўздан фойдаланилди. “Сўз” дейилганда, сўз қўлланишини (словоупотребление) назарда тутилди, яъни сўзнинг материал қисми (лексик маъно) билан формал (идеал) қисми (грамматик маъно) бир бутунликда - яхлитликда тушунилди.

Лугат сўзлиги қилиб сўзшакл (сўзформа) қабул қилинди. Сўзшакл термини, айниқса, структурал ёки математик лингвистикада кенг қўлланоқда. Матн ичида икки бўшлиқ (пробел) орасидаги бир бутунлик (“ҳарфлар занжири”) - сўзформаси ҳисобланади.

Олинган 100000 сўз ўн бўлак (бештаси - прозаик; бештаси - поэтик)ка бўлиниб, 10000 танланмадаги сўзнинг ҳар бири алоҳида-алоҳида ўзидан олдингиси билан қиёсан ўрганилди. Бундан мақсад, биринчидан, ҳар бир ўн минг янги сўз (лугат бирлиги – сўзшакллари)нинг миқдори ва умумий 100000 даги ўртача танланма частотасини аниқлашдир. Статистикада ўрта танланма частота (\bar{x}) қуйидагича аниқланади.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

бунда $x_i (i = \overline{1, n})$ - танланма частота; n - кузатилган танланма сони.

Масалан;

$$\bar{X} = \frac{4622 + 4826 + 4260 + \dots + 5158}{10} = \frac{49232}{10} = 4923,2/127,1^1$$

Олинган ўн танланма бўйича $\bar{x} = 4923,2/127,1$ (4174/84 + 5173/201). Бундан прозада $\bar{x} = 4588,4/117,6$ (4174/84 + 4742/159); поэзияда $\bar{x} = 5258/186,6$ (5037/105 + 5173/201).

Ўтган асрнинг бошлари даври тили лексикаси юзасидан намуна сифатида Х.Х.Ниёзий “Девон”и устида лингвостатистик тадқиқ олиб борилди ва конкорданси тузилди.

Х.Х.Ниёзий асарлари тили ва услуби ўзбек адабий тилининг шаклланиши ва равноқ топишида, унинг ҳозирги кундаги камолотга эришувида муҳим ўрин тутди.

Адибнинг Ўзбекистон Фанлар академияси Алишер Навоий номидаги Тил ва адабиёти институтида тайёрланган, “Фан” нашриёти томонидан 1979-1981 йилларда нашр этилган тўрт жилдлик “Мукаммал асарлар тўплами”даги “Девони

¹ Қия чиқкандан олдинги рақам умумий танланмадаги ўртача частота ҳисобланади, яъни \bar{X} жамғиға нисбатан бўлиб, бу ерда мустақил ва ёрдамчи сўзлар ҳам, ономастика ва топонимлар ҳам биргаликда ҳисобга олинган. Дроблан кейинги рақам эса дроблан олдинги рақам ичидagi ономастика ва топонимларнинг алоҳида ўртача танланма частотасидир. Куринадикси, умумий танланмада $\bar{X} = 4923,2$ бўлиб, бунинг ичида ономастика ва топонимларнинг ўрта танланма частотаси - $X = 127,1$ деб тушуналиши керак.

асосида “Ҳамза Ҳакимзода Ниёзийнинг девони тили лексикасининг статистик структураси, частотали лугат ва конкорданс” яратилди.

Ҳ.Ҳ.Ниёзий “Девон”и тилининг частотали лугати ва конкорданси структура ва мазмун-мундарижа эътибори билан умумхалқ тили ва адабий тил лугатларидан фарқ қилади.

Ҳ.Ҳ.Ниёзийнинг “Девони”и 1905-1914 йилларда ёзилган шеърларидан ташкил топган бўлиб, унда 177 та шеър келтирилган. Шулардан 165 таси ўзбекча, 12 таси форс-тожикча; 177 та шеърдан 150 таси ғазал, қолганлари мураббаъ, мухаммас ва мусаддас.

1979 йилда chop этилган “Девон”да эса 139 та шеър келтирилган бўлиб, унинг 123 таси ғазал, 3 таси мураббаъ, 12 таси мухаммас ва биттаси мусаддас; 139 та шеърдан 128 таси ўзбекча, қолган Ҳн биттаси форс-тожикчадир.

Шоир “Девон”га кирган шеърларнинг асосий мазмун-мундарижасини ҳаёт ва ҳаёт гўзаллигини тараннум этиш, маърифатпарварлик ва инсонийликни кадрлаш, улуғлаш билан бирга, ўша давр адабий муҳит қусурлари, жаҳолат ҳамда нодонликларни қоралаш, танқид қилиш ташкил этади. “Девон”даги барча мустақил сўз(сўзшакл)лар 6417 та бўлиб, булар матнда 64807 марта қўлланган. Мустақил ва ёрдамчи сўз(сўзшакл)лар сони 6325 та. Улар 16555 марта ишлатилган, умумий матн ҳажмининг 98,5 фоизини ташкил этади. Киши ва географик номларнинг умумий матндаги салмоғи эса 1,5 фоиз (92 та сўзшакл 252 марта қўлланган)дир.

“Девон” матнида қўлланган сўзларнинг статистик тақсимланиши қуйидагича: 1 марта қўлланган сўзларнинг сони — 3648 (65%); 2 марта қўлланган сўзлар — 782 (14%); 3—4 мартадан қўлланган сўзлар — 558 (10%); 5-9 мартадан қўлланган сўзлар — 394 (7%); 10 ва ундан ортик қўлланган сўзлар сони 252 (4%) дир. Ҳн битта форсча-тожикча шеърларида: 1 мартадан қўлланган сўзлар — 649 (45,636%); 2 мартадан қўлланган сўзлар — 166 (13,799%); 3-4 мартадан қўлланган сўзлар — 98 (8,146%); 5-9 мартадан қўлланган сўзлар — 92 (7,648%); 10 ва ундан ортик марта қўлланган сўзлар сони - 298 (24,771%) тадир.

Диссертацияда Ҳ.Ҳ.Ниёзий “Девон”и тили (ўзбекча, форсча-тожикча) лексикаси (сўзшакллари)нинг лингвостатистик таҳлили юзасидан яратилган статистик структураси, шунингдек, алфавитли - частотали лугати ва конкорданси асосида “бор” сўзи мисолида намуна келтирилди. Ана шулар асосида “Девон”да қўлланилган сўзлар ва уларнинг маъно оттенкалари, фонетик шакли, услубий хусусиятлари орқали Ҳамза Ҳакимзода Ниёзийнинг сўз қўллашдаги маҳоратини ҳам мукаммал ўрганилиши мумкин.

3. ХУЛОСА

1. Кирилл ёзувига асосланган ўзбек алифбосидаги ҳарфларнинг частотасини лингвостатистик тадқиқ этиш тилининг муҳим статистик тавсифини берди. Лингвостатистик тадқиқ ва таҳлил шуни кўрсатдики, ўзбек алифбоси сирасидаги а, и, н, л, р ҳарфлари умумий матн ҳажмининг қарийб ярми(50 %)ни; о, д, б, т, г, м

харфлари эса 25 % ини ташкил этади. Демак, а, и, н, л, р, о, д, б, т, г, м харфлари умумий матн ҳажмининг 75 % га яқинини ташкил этди. Проза, поэзия, ижтимоий-гуманитар ва аниқ фанлар каби соҳаларнинг ўзаро тафовутга олиб келувчи харфлари б, к, е, ф, ш, м, ц лардир. Харфлар частотасининг лингвостатистик структураси ўзбек алифбоси билан боғлиқ барча ахборот технологияларида техник жиҳозларни, телестайп ва машина таржимаси учун кодировкаларни оптимал ўрнатишда, келгусида янги ўзбек компьюттери модели(тизими)нинг яратилишида юкори частотали ва фаол харфлар ҳамда икки харфли бирикмалар асосида клавиатурадаги тутмачаларни энг қулай жойлаштириш имконини беради. Шунингдек, ўзбек алифбоси харфларини саккизлик система асосида эффектив кодлаштириш учун алифбонинг “янги тартиби”ни ҳам тузиш мумкин.

2. Дедуктив баҳо беришнинг ўлчов воситаларидан бўлган энтропиянинг статистик, методда аниқланиши тилнинг график ва фонологик системасини тадқиқ этишда фойдали натижалар беради. Шунингдек, матн энтропиясининг аниқ баҳосини белгилаш математик ва амалий тилшуносликнинг долзарб масалаларидан бири ҳисобланади. Ўзбек тилининг босма матнлари энтропиясининг аниқланиши шу тил харфлари билан боғлиқ ахборотларни узатиш тезлигини оширишда аҳамиятлидир. Лингвостатистик тадқиқот натижасида ўзбек тилининг бир харфи 5,17 “бит” ёки $36=2^{5,17}$ информация берди. Харфлар энтропияси ёрдамида ўзбек тилидаги сўзларнинг харфлар бўйича мезони (чегараси) аниқланди. Жумладан, ўзбек тилидаги сўзларни аниқлаш учун 5-6 гача харфдан фойдаланилиши исботланди.

3. Ўзбек тили фонемалар системасини унли ва ундошларга ажратиб, ўзаро фоиц нисбатларини солиштириш, унлилар системаси (ўзбек тилидаги барча фонемаларнинг 19 % и) умумий матн ҳажмининг 42 % и (аниқроғи 41,751896; $F = 72012$)ни, ундошлар эса 58 % и (аниқроғи 58,248104; $F = 100464$)ни ташкил этишини кўрсатди. Ҳозирги ўзбек тилида фонология нуқтаи назаридан (\bar{x}) 100 унлига 140 ундош қўлланади, яъни ўзбек тили учун унли ва ундошларнинг ўзаро миқдорий муносабати жиҳатидан $N_c : N_v = 1,40$ (1,3951008) статистик параметри ҳисобланади. Тилдаги олти унли (а, о, и, у, э, ў) ва 6 сонор (н, л, р, м, й, н) нинг умумий матндаги салмоғи 68 % (67,776387). Демак, ҳозирги ўзбек тилини фонематик таркиби жиҳатидан сонангли деб, тавсифлаш мумкин. Шунингдек, ўзбек тилининг проза матнларидаги “фонематик занжир”да унли ва ундошларнинг ўзаро миқдорий муносабатлари - консонангли коэффициенти - $N_c : N_v = 1,37$ (1,366942); поэзия матнларида - $N_c : N_v \approx 1,47$ (1,470742); ижтимоий-гуманитар матнларида - $N_c : N_v \approx 1,37$ (1,372656) ва аниқ фанлар матнларида - $N_c : N_v \approx 1,37$ (1,374227). Умуман, ўртача (\bar{x}) - $N_c : N_v \approx 1,40$ (1,3951008).

4. Ҳозирги ўзбек тилидаги 6 унли фонема проза матнларида қўлланган ҳамма фонеманинг 42 % ини; 25 ундош фонема эса 58% ини ташкил этади. Абсолют частоталар муносабати ($\Sigma = 42960$ га нисбатан) $V18150 : C24810$. Бундай матнлардаги “фонематик занжир”да ўртача 100 унлига ≈ 137 ундош қўлланади, яъни консонангли коэффициент - $N_c : N_v \approx 1,37$ дир.

Поэзия матнлари бўйича барча унлиларнинг умумий матндаги салмоғи 40 %, ундошларники 60 % дир. Абсолют частоталар муносабати ($\Sigma = 42603$ га нисбатан) $V_{17243} : C_{25360}$.

Поэзия матнларидаги “фонематик занжир”да унли ва ундошларнинг нисбий частота миқдорий муносабати (консонантли коэффицентги) - $Nc : Nv \approx 1,47$ дир.

Ижтимоий-гуманитар матнларида Ўзбек тилидаги барча унлиларнинг умумий матндаги салмоғи 42 %; ундошларники 58% дир. Абсолют частоталар муносабати ($\Sigma = 43524$ га нисбатан) $V_{18344} : C_{25180}$.

Бундай матнлардаги консонантли коэффицент - $Nc : Nv = 1,37$.

Ўзбек тилида ундошларнинг тилдаги барча фонема (VC) лар частотасига нисбатан умумий матндаги салмоғи 58 %ни ташкил этади (Проза, ижтимоий-гуманитар ва аниқ фанлар матнларининг ҳар бирида – 58 % дан; поэзия матнларида эса 60 %). Бундан сонорларнинг салмоғи – 26 % (Поэзия ва аниқ фанлар матнларида – 27 % дан; ижтимоий-гуманитар матнларида – 26 %; проза матнларида – 25 %).

5. Ўзбек тили фонемаларининг сўздаги позициясига кўра функционал “нагрузка”си турличадир:

а) унлилар ва сонорлар сўз бошидагига нисбатан сўз охирида, сўз боши ва сўз охиридагига нисбатан сўз ўртасида энг кўп ва фаол қўлланиши билан характерланади (икки унлили бирикмалар ҳам шундай);

б) портловчи ва сирғалувчи ундош фонемалар сўз охиридагига нисбатан сўз бошида, сўз боши ва сўз охиридагига нисбатан сўз ўртасида энг кўп ва фаол қўлланиши билан характерланади.

6. Ўзбек тилида V, C ва “А”нинг ўзаро жуфт ҳолда қўллана олиш миқдори: $VV < \Delta V$; $VV < V\Delta$; $CC < \Delta C$; $CC < C\Delta$; $VC > VV$; $VC > CC$; $CV > VV$; $CV > CC$.

Умуман, ўзбек тилидаги “фонематик занжир”да жуфт ҳолда энг кўп қўлланидиган структурали тип CV дир.

7. Кузатилган матнларда ўзбек тилида назарий жиҳатдан қўлланиши мумкин бўлган барча иккифонемали бирикмаларнинг 74%и учради. Ўзбек тилидаги ҳар бир фонема тилдаги барча фонемаларнинг неча фоизи билан жуфт ҳолда кўпроқ қўллана олиши қуйидагича белгиланади: “И” фонемаси ҳамма фонемаларнинг 97 %; “А”, “Н”, “Р”, “О” ларнинг ҳар бири ҳамма фонемаларнинг 94 %; “Ј”, “У”, “Е(Э)” ларнинг ҳар бири ҳамма фонемаларнинг 90 %; “Й” - ҳамма фонемаларнинг 87 %; “Ш” - ҳамма фонемаларнинг 84 %; “М”, “К”, “З”ларнинг ҳар бири ҳамма фонемаларнинг 81 % и билан қўллана олади.

8. Икки фонемали бирикмалар таркибидаги унли ва ундош фонемаларнинг қўлланиш даражаси турлича характердадир: Н, Р, Л, Г, Д, Т, М, К, З, С ундошларининг “А” ва “И” унлиси билан ёнма-ён келган бирикмалари (CV ёки VC) ўта фаол ҳисобланади.

Икки фонемали (уч фонемали, тўрт фонемали ва х.к.) бирикмалар ҳақидаги лингвостатистик параметрлар ўзбек тилининг фонемалар дистрибуциясини ёки умуман, фонологик системасини кенгроқ тасвирлашда катта аҳамиятта эга.

Бундан, албатта, фонология даражасида синтагматик бирликнинг статистик моделини тузишда фойдаланилади.

9. Ўзбек адабий тилининг оригинал проза матнларида бўгин типларининг дастлабки туртгаси (CV, CVC, V, VC) кўпгина туркий (Масалан: озарбайжон, қорақалпоқ, нўғай) тиллардагидек асосий салмоқни эгаллайди. Улар умумий матн ҳажмининг 99% (аниқроғи 98,657 %) ини ташкил этган. Ўзбек тилининг бадий-прозаик стилида бўгинларнинг 90 % ундош фонема билан, қолган 10 % эса унли фонема билан бошланади.

Бўгиннинг CV типи бошқа барча типларга нисбатан энг кўп қўлланиши билан характерланади. CVнинг умумий мантдаги салмоғи - 48,983 % дир. Бўгиннинг CV ва CVC типи ўзбек тилидаги сўзлар структурасини ташкил этишда фаол қўлланади.

Бўгиннинг V, VC, CV типлари сўз бошида (сўз охирида ёки сўз ўртасида эса бўгиннинг CVC типи) энг кўп қўлланиши билан характерланади. Ўзбек тилининг оригинал бадий-прозаик стилига икки ва уч фонемали бўгин типи хосдир. Уларнинг умумий матидаги салмоғи 93,397 %.

Бўгин типларини ҳосил қилишда P, Л, Н, Й фонемалари фаоллиги билан характерланади. Олинган таъланма ҳажми доирасида 1294 хил “бўгин хили” аниқланиб, булардан 58 хили (F = 106 ва ундан юқори) умумий мати ҳажмининг яримдан кўпроғини - 56,386667 % ини ташкил этади.

10. Лингвостатистик кузатиш шуни кўрсатдики, ўзбек тилининг оригинал бадий-прозаик стилида сўзнинг ўртача узунлиги 2,572 бўгин сонига тенг. Тидаги очик бўгинларнинг ёпиқ бўгинларга нисбатан миқдорий муносабати - 54,773 : 45,227. Ўзбек тилининг оригинал бадий-прозаик стилида бўгин таркибидаги унли (V) ва ундош (C)ларнинг ўзаро миқдорий муносабати - 1,37 (N_c : N_v = 1,367167).

Ўзбек тилида бўгиннинг CCV, CCCV, CCVC, VCCC, VCCCC, CVCCC, CCVCC, CCCVC, CCVCCC типлари рус тили ва у орқали ўзлашган сўзлар таъсири маҳсулидир.

11. Ҳозирги ўзбек тилида назарий жиҳатдан морфемаларнинг ўрилашувида, асосан, куйидаги модель етакчидир:

асос (“аснова” маъносида) + лар + им + нинг

Бироқ буларнинг лингвостатистик параметрлари ҳозирча тўлиқ очиқгани йўқ. Статистик параметрлар асосида уларнинг лингвостатистик тадқиқ ва таҳлил этилиши мақсадга мувофиқдир. Сўзшакллари частотали лугати асосида сўз ёки сўзшакллари морфемаларга ажратишда ҳозиргача амалда қўлланиб келаётган тартибдан фарқли ўлароқ чап томондан (асос томондан) ўнг томонга қараб эмас, балки ўнг томондан — сўзшаклидаги охириги морфемадан чап томонга қараб бўлиб келиниши, яъни морфемаларга ажратилиши мақсадга мувофиқдир. Бу эса морфемага берилган умумий таърифга тўла мос келади. Юқоридаги ҳолатларнинг барчаси турли функционал стиль доирасида сўзшакл базасида лингвостатистик таҳлил этилишини талаб этади. Ана шундагина лингвостатистик аспектда сўзшакллардаги морфем таҳлил муаммоси илмий-назарий ва амалий жиҳатдан ўз

счмийни топипи, шак-шубҳасиздир. Шунингдек, ўзбек тили сўзшаклларида морфемаларнинг автоматик тақсимланиши муаммоси кейинчалик ҳам ўз долзарблигини сақлаб қолади.

Ўзбек тилида мустақил сўз туркумлари морфем структурасининг алоҳида-алоҳида лингвостатистик ва қиёсий-типологик аспектда тадқиқ этилиши назарий ва амалий жиҳатдан муҳимдир.

Тил структурасининг энг кичик бўлинмас маъноли қисмлари – морфема, монема, темемалар тавсифининг лингвостатистик тадқиқ этилиши долзарблиги билан ажралиб туради. Шунингдек, ўзбек тили матнларидаги морфемалар сонининг сўзшакллари сонига муносабати тил синтезининг ўлчови (синтетиклик индекси) бўлиб хизмат қилади. Чунончи, вьетнам тилида сўзлар (сўзшакллар) деярли бир морфемали бўлганлиги боис синтетиклик индекси – 1,06. Эскимос тилида эса 3,72 дпр. Ушбу кўрсаткич инглиз тилида – 1,68, рус тилида – 1,90, украин тилида – 1,80, ўзбек тили эса тахминан 1,50 – 1,60 ораллиғидадир.

12. Ўзбек болалар адабиёти материаллари асосида 100000 сўзлик олинма ҳажмида бир марта қўлланган сўзшакллар умумий матн ҳажмининг 47,90% ини; икки марта қўлланган сўзшакллар умумий матн ҳажмининг 17,24% ини; 3-4 марта қўлланганлари – 13,32% ини; 5-9 марта қўлланганлари – 11,42% ини; 10 ва ундан зиёд қўлланганлари эса 10,12% ини ташкил этади.

Болалар адабиёти тилида қўлланган сўзшакллариининг сўз туркумлари бўйича статистик тақсимланишида асосий салмоқ от, феъл ва бошқа туркумларга тушади.

Ҳозирги ўзбек болалар адабиёти стилида от туркумининг кичрайтириш формалари *-ча, -чоқ (-чак)* аффикслари, шунингдек, от туркумининг эркалаш формасини ифодаловчи *-жон, -вой, -хон, -ой* аффикслари ҳамда ўхшатиш формасини билдирувчи *-дай (-дек)* кўп ва фаол қўлланиши билан характерланади. Ўзбек болалар адабиёти материаллари асосида юз минг сўз қўлланиши олинма ҳажмида аниқланган сўзшакллариининг 4/5 қисмига яқини этимологик жиҳатдан ўзбекча(туркийча) сўзлардир. Қолганларини тожикча-форсча, тожикча-ўзбекча, тожикча, ўзбекча, арабча, арабча-ўзбекча, русча, русча-ўзбекча сўзлар ташкил этади.

13. Ҳозиргача ўзбек тилида турли функционал стиллар доирасида частотали лугатлар старли танланма материал(матн)лар асосида тузилганча йўқ. Ҳозирги ўзбек тили лексикаси лингвостатистик тадқиқ этилса, унинг лингвистик банки (машина фонди) асосида лексикограмматик аспектда лингвостатистик параметрлар яратилади. Булар эса ўз навбатида, биринчидан, бир тилдан иккинчи тилга автоматик таржима соҳалари (машина таржимаси)да, иккинчидан, бундай лексикостатистик тадқиқот натижаларидан таълим жараёнида, ўқиш-ўқитиш ишларида, дарслик ва ўқув қўлланмалари яратишда, илму лугатлари, лугат-минимумлар, изоҳли, этимологик ва кўп тилли лугатлар тузишда ҳамда қиёсий ва квантитатив-типологик ва информацион-статистик тадқиқотлар олиб боришда манба (асос) бўлиб хизмат қилиши мумкин.

4. ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ

Монография ва илмий журналларда эълон қилинган мақолалар

1. Ризаев С.А. Ҳозирги ўзбек адабий тилида ҳарф ва фонемаларнинг қўлланиш частотаси (Тилшунослик, адабиётшунослик ва физика материаллари бўйича). – Т.: Фан, 1967. - 20 б.
2. Ризаев С.А. Тилшунослиқда математик методнинг қўлланишига доир. – Т.: Билим, 1973. - 23 б.
3. Ризаев С.А. Бўгин структурасини статистик ўрганиш тажрибасидан. – Т.: Фан, 1975. - 172 б.
4. Ризаев С.А. Кибернетика ва тилшунослик. – Т.: Ўзбекистон, 1976. - 16 б.
5. Ризаев С.А. Ўзбек болалар адабиёти тилининг частотали луғати. – Т.: Фан, 1980. - 160 б.
6. Ризаев С.А., Бўронов Н. Абдулла Қаҳҳор “Синчалак” повести тилининг частотали луғати. – Т.: Ўқитувчи, 1986. - 172 б.
7. Ризаев С.А. Статистическая структура лексики языка “Дивана” Хамзы Хаким-заде Ниязи (Частотный словарь и конкорданс). -Т.: Фан, 1989. -248 с.
8. Ризаев С.А., Бўронов Н. Имло луғати (Бошланғич синфлар учун). -Т.: Ўқитувчи, 1992. – 368 б.
9. Ризаев С.А. Имло ва талаффуз луғати (Бошланғич синф ўқувчилари учун). -Т.: Ўқитувчи, 1998. - 80 б.
10. Ризаев С.А. Имло луғати. “Х”ми ёки “Х”? - Т.: Кўҳи-Нур, 2002. -52 б.
11. Ризаев С.А., Абдурахмонов М. Ўзбек тилшунослигининг лингвостатистика соҳасига доир. - Т., 2004. - 48 б..
12. Ризаев С.А. Ўзбек тилининг имло луғати. -Т.: Кўҳи-Нур, 2005. -160 б.
13. Ризаев С.А. Ўзбек тилшунослигида лингвостатистика муаммолари. -Т.: Фан, 2006. -296 б.
14. Ризаев С.А. Имло луғати. “Х”ми ёки “Х”? (Тузатилган ва тўлдирилган наشري) - Т.: Ўзбекистон миллий кутубхонаси нашриёти, 2008. -60 б.
15. Ризаев С.А. Ўзбек адабий тилининг фонологик системасини статистик текшириш ҳақида // Ўзбек тили ва адабиёти. –Тошкент, 1970. - 1-сон. – Б. 71-75.
16. Ризаев С.А. Статистическое исследование частотности букв в письменном тексте современного узбекского языка. // Общественные науки в Узбекистане. – Ташкент, 1970.- N 10. – С. 65-68.
17. Ризаев С.А. Ҳарфлар баҳси // Мухбир. – Тошкент, 1971. -9-сон. -Б. 46-47.
18. Ризаев С.А. Из опыта исследования статистической структуры слога в узбекском языке // Общественные науки в Узбекистане. –Ташкент, 1972. -№ 1. - С.72.
19. Ризаев С.А. Ўзбек тили унли фонемаларини статистик метода ўрганиш // Ўзбек тили ва адабиёти. – Тошкент, 1972. -4-сон. – Б. 35-39.
20. Ризаев С.А. Кибернетика шарофати // Фан ва турмуш. –Тошкент, 1973. - 2-сон. – Б. 12-13.

21. Ризаев С.А. Ундошлар классификациясига доир бир масала // Ўзбек тили ва адабиёти. – Тошкент, 1974. -1-сон. – Б. 77-79.
22. Ризаев С.А., Рўзикулов Э. Испан тили дарслиги лексикасини статистик текшириш тажрибаси // Совет мактаби. –Тошкент, 1975. - 10-сон. – Б. 73-74.
23. Ризаев С.А. Бўгин структурасини статистик ўрганиш // Ўзбек тили ва адабиёти. – Тошкент, 1977. - 3-сон. – Б. 62-64.
24. Ризаев С.А. Болалар адабиёти тилида сўз қўллашни статистик тадқиқ этиш ҳақида // Ўзбек тили ва адабиёти. – Тошкент, 1985. - 5-сон. - Б. 44-47.
25. Ризаев С.А., Рўзикулов Э. Газета тили лексикаси ва статистик тадқиқ // Совет мактаби. –Тошкент, 1987. - 1-сон. – Б. 17.
26. Ризаев С.А. “Синчалак” повести тили лексикасининг статистик тадқиқи // Совет мактаби. –Тошкент, 1987. - 8-сон. – Б. 23-24.
27. Ризаев С.А. Тилшуносликда замонавий ахборот технологияларининг қўлланишига доир // Ўзбек тили ва адабиёти. – Тошкент, 2006. - 5-сон. -Б.95-97.
28. Ризаев С.А. “Ҳ.Ҳ.Нижзий “Девон”и тили лексикасининг лингвостатистик тузилиши ҳақида. // Тил ва адабиёт таълими. –Тошкент, 2006. -6-сон. -Б.57-59.
29. Ризаев С.А. Лингвостатистик тадқиқотларига доир. // Ўзбек тили ва адабиёти. – Тошкент, 2007. -6-сон. -Б. 76-79.

Илмий тўпламларда чоп этилган мақола ва тезислар

30. Ризаев С.А. Ҳозирги ўзбек адабий тили лексикасини статистик текшириш ҳақида баъзи маълумотлар // Тўплам: Вопросы кибернетики и вычислительной математики. – 23- сон. - Т.: Фан, 1968. – Б. 37-63.
31. Ризаев С.А. Ўзбек тилининг ёзма нутқида ҳарфларнинг тақсимланиши ҳақида эмпирик маълумотлар // Тўплам: Ўзбек тилшунослигининг баъзи масалалари. – Т.: Фан, 1969. – Б.4-6.
32. Ризаев С.А., Фозилов Э., Ёқубова Н., Чичулина Л. Применение ЭВМ при исследовании статистики фонем и слов узбекского языка // Тўплам: Автоматическая переработка текста методами прикладной лингвистики. - Кишинев, 1971. - С. 64 - 65.
33. Ризаев С.А., Бектаев Қ.Б. О критерии принадлежности диалектов к одному языку // Тўплам: Статистика казахского текста. Часть 1. Труды группы Статистико-лингвистическое исследования и автоматизация. - N 3. - Алма-Ата, 1973.- С. 643-658.
34. Ризаев С.А., Фуломов А.Г. Опыт лингвостатистического исследования структуры слова в современном узбекском языке // Советская тюркология в развитие тюркских языков в СССР. - Алма-Ата, 1976. - С. 312.
35. Ризаев С.А., Юсупова Д. О дистрибутивно-статистическом исследовании лексико-морфологической структуры слова современного узбекского языка // Тўплам: Исследование по литературоведению и языкознанию. – Т.: Фан, 1977. – С.198-203.

36. Ризаев С.А. Дистрибутивно-статистический анализ структуры слова в современном узбекском языке // Тўплам: - Школа-семинар по прикладной и инженерной лингвистики. - Махачкала, 1978. – С. 38 - 39.

37. Ризаев С.А., Аълоев Й., Рахмонова С.А., Рўзикулов Э., Караева М.Х. О методике и принципах составления частотных двуязычных учебных словарей // Школа-семинар по оптимизации преподавания иностранных языков с помощью технических средств. - Кишинев, 1979. – С. 59.

38. Ризаев С.А., Бектаев Қ.Б., Джубанов А.Х., Ахабобов А. Состояние и перспективы исследования тюркских языков с применением статистико-инженерных методов // Тўплам: Языкознание (III Халқаро туркологик анжуман). - Т.: Фан, 1980. – С. 168-169.

39. Ризаев С.А. Статистико-вероятностный анализ структуры слова в современном узбекском языке (на фонемном и морфемном уровнях) // Тўплам: Языкознание (III Халқаро туркологик анжуман). - Т.: Фан, 1980.- С. 179-180.

40. Ризаев С.А. Статистический анализ структуры слова в современном узбекском языке // Тўплам: Статистическая оптимизация преподавания языков и инженерная лингвистика. - Чимкент, 1980. – С. 280 - 281.

41. Ризаев С.А. К вопросу частотности двухфонемных сочетаний в современном узбекском литературном языке // Тўплам: Опыт экспериментального и структурного изучения языка. - Т.: Фан, 1982. – С. 45-52.

42. Ризаев С.А., Ачилов Т.А. Компьютерные программы по грамматике немецкого и узбекского языков // Материалы семинара “Статистическая оптимизация преподавания языков и компьютеризация обучения”. Часть II. - Чимкент-Минск, 1990. - С. 206 - 207.

43. Ризаев С.А., Тайлокова М., Ҳазраткулов А., Бўронов Н. Ўзбек тилининг алфавитли - частотали лугати ҳақида // Тўплам: Ўзбек филологияси масалалари. - Т.: Фан, 1990. – Б.189-193.

44. Ризаев С.А. Статистика ва она тили ўқитиш // Ўзбек тилини ўрганиш ва ўргатишни қайта қуриш ҳамда такомиллаштиришнинг долзарб масалалари. Ўзбек тили доимий анжумани Биринчи йиғини маърузаларининг тезислари (Самарқанд, 1991, 7-10 январь). - Самарқанд, 1991. – Б.112-113.

45. Ризаев С.А. Русча тилшунослик атамаларининг ўзбекча муқобилига доир // Ўзбек тилида илмий-техникавий адабиётларни ёзиш, таржима ва нашр қилиш муаммоларига бағишланган Республика илмий-техникавий анжуманида қилинган маърузаларнинг қисқача баёни (Т., 1991, 10-12 июль). - Т., 1991. – Б.18-19.

46. Ризаев С.А. Ўзбек тилшунослигининг айрим назарий ва амалий муаммолари // Абдурауф Фитрат миллий феномени, адабий, илмий-назарий мероси, ижтимоий-сиёсий фаолиятига бағишланган Биринчи Республика илмий-амалий анжуман Материаллари (Бухоро, 1992, 21-23 май). - Бухоро, 1992. –Б. 97-99.

47. Ризаев С.А. Луғат устида ишлашга доир // Республика Халқ таълими тизими муассасаларидаги тажриба-текширув ишларининг самарадорлигини опирриш юзасидан илмий-амалий педагогик кенгаш маърузалари Тезислари. - Т.: ЎзРПФИТИ, 1993. -Б. 25-28.

48. Ризаев С.А. О банке терминологических данных узбекского языка (На примере узбекских лингвистических терминов) // Международный Казахско-Турецкий университет имени Х.А.Яссави. Материалы Международной конференции “Компьютерный фонд терминов тюркских языков” (Чимкент, 1995, 23-25 ноября). - Туркистан-Чимкент, 1995. - С. 130-131.

49. Ризаев С.А. Ўзбек тили фонемалари қўлланиш частотасига доир // Мирзо Улугбек номидаги Ўзбекистон Миллий университетининг “Тилшуносликнинг долзарб масалалари” мавзусидаги Илмий мақолалар тўплами. - Т.: Университет, 2004. - Б.35-37.

50. Ризаев С.А. Тилшунослик атамаларига бир назар // ТДИУ нинг “Иктисодий муносабатларни ўрганишда иктисодий атамаларидан тўғри фойдаланиш зарурияти ва муаммолари” мавзусидаги Республика илмий-амалий анжуманининг Маърузалар тезислари тўплами (2004 йил, 18 май). - Т., 2004. -Б.81-82.

51. Ризаев С.А. Миллий маданиятимиз белгиси – тил (Давлат тилига муносабат муаммолари // ТДЮИ нинг Тил ҳуқуқи – эл ҳуқуқи “Давлат тилининг ҳуқуқий мақоми: муаммолар, ечимлар” мавзусида ўтказилган илмий-амалий анжуман материаллар (II сон). - Т., 2005. -Б.15-19.

52. Ризаев С.А. Тилшунослик ва кибернетика // Ўзбек тили таълими жараёнида фанлараро боғланиш масалалари мавзусидаги “Ўзбек тили” доимий анжумани саккизинчи йигинининг материаллари (Урганч ш., 2005 йил, 21-22 апрель). - Т., 2005. -Б.154.

53. Ризаев С.А. Морфемика. Морфемаларнинг статистик структурасига доир // ЎзМУ: “Тилшуносликнинг долзарб масалалари” (Илмий мақолалар тўплами) III. -Т., 2006. -Б.56-62.

54. Ризаев С.А. Ўзбек тилшунослигида лексикостатистик тадқиқотга доир. Тўплам (ТИМИ). -Т., 2007. -Б.11-15.

55. Ризаев С.А. Ўзбек лексикографиясининг тараққиётига доир (луғатлар муҳим қўлланмадир) // XXI асрда ўзбек тили таълими масалалари “Ўзбек тили” доимий анжумани IX йигинининг материаллари (Термиз ш., 2007 йил 19-20 апрель). -Т., 2007. -Б. 263-265.

56. Ризаев С.А., Боймирзаева С. Математик ва компьютер тилшунослиги муаммолари олим талқнида. СамДУ. - “Услубшунослик ва фразеологиянинг долзарб муаммолари” мавзуси бўйича профессор Бекмурод Йўлдашев таваллудининг 60 йиллигига бағишлаб ўтказилган илмий-назарий конференция (Самарқанд ш., 2007 йил, 7-8 сентябрь) Материаллари. -Самарқанд, 2007. - Б. 32-34.

Филология фанлари доктори илмий даражасига талабгор Ризаев Суннатулла Абдуманнаповичнинг 10.02.01 – Ўзбек тили ихтисослиги бўйича “Ўзбек тилининг лингвостатистик тадқиқи” мавзусидаги диссертациясининг

РЕЗЮМЕ СИ

Таянч сўзлар: математик лингвистика, компьютер лингвистикаси, лингвостатистика, тақсимот қонуни, кодлаш, микдорий таҳлил, вариация коэффициенти, фойдаланиш коэффициенти, корреляция коэффициенти, структурал лингвистика, амалий лингвистика, ҳарфий статистика, фонологик статистика, бўғин статистикаси, морфем статистика, лексик статистика, статистик қонуният (параметр), лингвистик банк, машина фонди.

Тадқиқот объектлари: Ўзбек тилининг турли функционал стиллари оригинал материаллари (проза, поэзия, ижтимоий-гуманитар адабиётлар: публицистика, философия, тарих ва аниқ фанлар: физика ва математика), шунингдек, Абдулла Қаҳҳорнинг “Синчалак” повести тили лексикаси ҳамда Ҳамза Ҳакимзода Ниязий “Девони” тили лексикасининг статистик структураси (частотали лугат ва конкорданс).

Ишнинг мақсади: Ўзбек тили босма матнларида ҳарфлар, фонемалар, бўғинлар, морфемалар ва лексемаларнинг қўлланиш частоталари юзасидан статистик тадқиқот натижасида лингвостатистик қонуният (параметр)ларни аниқлаш.

Тадқиқот методлари: статистик, индуктив ва дедуктив, киёсий-типологик, систем-структур, информацион-статистик ва аналитик-синтетик таҳлил, моделлаштириш ва б.

Олинган натижалар ва уларнинг янгилиги: тил ва ЭҲМ асосида лингвостатистик қонуният (параметр)лар ва тавсифлар аниқланди ҳамда статистик ва информацион-типологик натижаларга эришилди. Тилшунос, математик ва дастурчи (программист)ларнинг ҳамкорлигида Ўзбек тили лингвистик банкларини, машина фондини, шунингдек, “одам-машина-одам” коммуникатив алоқасини ўрнатиш имконини тақозо этди.

Амалий аҳамияти: иш натижаларидан умумий тилшунослик назарияси (хусусан, лингвостатистика) бўйича махсус курслар, семинарлар ташкил этишда, техника жиҳозлари (ускуналари) клавишларини қулай ва оптимал жойлаштиришда, дарсликлар, ўқув қўлланмалари ва турли типдаги лугатлар ва лугат-минимумларини тузишда, шунингдек, компьютер лингвистикасида кенг фойдаланиш мумкин.

Татбиқ этиш даражаси ва иқтисодий самарадорлиги: муаллифнинг тавсиялари Ўзбек тилшунослиги лингвостатистикасининг шаклланиши ва тараққиётида амалий қўлланма сифатида фойдаланилади. Шунингдек, тадқиқот натижаларидан Ўзбек лингвистик банкларининг (Ўзбек тили машина фондининг) яратилишида фойдаланиш мумкин.

Қўлланиш соҳаси: “Умумий тилшунослик”, “Структурал, амалий ва математик лингвистика”, “Ўзбек тилшунослиги” ва “Тилшунослик назарияси” ҳамда “Амалий математика”, “Компьютер лингвистикаси” фанларининг ўқитилишида, махсус курсларида қўлланилади.

РЕЗЮМЕ

диссертации Ризаева Суннатуллы Абдуманнаповича на тему “Лингвостатистическое исследование узбекского языка” на соискание ученой степени доктора филологических наук по специальности 10.02.01 - Узбекский язык.

Ключевые слова: математическая лингвистика, компьютерная лингвистика, лингвостатистика, закон распределения, кодирование, количественный анализ, коэффициент вариации, коэффициент оборачиваемости, коэффициент корреляции, структурная лингвистика, прикладная лингвистика, графическая статистика, фонологическая статистика, слоговая статистика, морфемная статистика, лексическая статистика, статистическая закономерность (параметр), лингвистический банк, машинный фонд.

Объекты исследования: оригинальные материалы разных функциональных стилей узбекского языка: художественная проза, поэзия, общественно-политическая литература (публицистика, философия, история), научная проза (физика, математика), лексика языка повести “Синчалак” Абдуллы Каххара, также лексика языка “Дивана” Хамзы Хаким-Заде Ниязи (частотный словарь и конкорданс).

Цель работы: определить лингвостатистическую закономерность (параметры) на основе лингвостатистического исследования частоты встречающихся букв, фонем, слогов, морфем и лексики на материалах печатных текстов узбекского языка.

Методы исследования: использованы методы статистический, индуктивный, дедуктивный, сравнительно-типологический, системно-структурный, информационно-статистический и аналитико-синтетический анализ, моделирование и другие.

Полученные результаты и их новизна: получены лингвостатистические параметры, разработаны лингвостатистические и информационно-коммуникационные технологии использования ЭВМ в узбекском языке, даны рекомендации по установке коммуникационных связей при создании лингвистических банков данных узбекского языка, при определении машинного фонда узбекского языка, а также разработана система “человек-машина-человек” на основе совместной координации усилий лингвистов, математиков и программистов.

Практическая значимость: результаты исследования могут быть широко использованы в теории общего языкознания (в частности, лингвостатистики) при организации специальных курсов лекций и семинаров, в создании технических средств обучения для удобного расположения клавиш в компьютерной клавиатуре, в создании учебников и учебных пособий, а также различных словарей. Практически неопенимы результаты исследования при использовании их в компьютерной лингвистике.

Степень внедрения и экономическая эффективность: рекомендации автора используются в качестве практического пособия по формированию и развитию лингвостатистики в узбекском языкознании. Результаты исследования могут быть также использованы в выработке узбекского лингвистического банка (машинный фонд узбекского языка).

Область применения: в учебных курсах “Общее языкознание”, “Структурная, прикладная и математическая лингвистика”, “Узбекское языкознание”, “Теория языкознания”, “Прикладная математика” и “Компьютерная лингвистика”.

RESUME

Thesis of Rizaev Sunnatulla Abdumannapovich on the scientific degree competition of the doctor of philological sciences on specialty 10.02.01 - The Uzbek language on the subject: "Linguostatistical investigation of the Uzbek Language"

Key words: mathematical linguistics, computer linguistics, linguostatistics, law of distribution, coding, quantitative analysis, factor of variation, factor of circulation, factor of correlation, structural linguistics, applied linguistics, graphic statistics, phonologic statistics, syllabic statistics, morphemic statistics, lexical statistics, statistical regularity (parameter), linguistic bank, machine fund.

Subjects of the inquiry: original materials of different functional styles of the Uzbek language (art prose, poetry, social-political literature (publicistic, philosophy, history), scientific prose (physics, mathematics). Besides, lexicon of language of Abdulla Kahhar's story "Sinchalak" and lexicon of language of "Devon" of Khamza Khakim-Zade Niyazi (frequency vocabulary and concordance).

Aim of the inquiry: to determine the linguostatistic law (parameters) on basis of linguostatistic research of frequency of letters, phonemes, syllables, morphemes, meeting, and lexicon on materials of printed texts of the modern Uzbek language.

Methods of the inquiry: statistical, inductive and deductive, comparative and typological, systemic and structural, informational, analytical and synthetic analysis, modeling were used.

The results achieved and their novelty: linguostatistic parameters were received, linguostatistic, information and communication technologies of computers usage in the Uzbek were elaborated, recommendations on installation of communication relations while creating the linguistic databanks of the Uzbek language and estimating the machine fund of the Uzbek language were worked out; also the system "person-machine-person" on the basis of the joint coordination of efforts of linguists, mathematicians and programmers were worked out.

Practical value: results of research can be widely used in the theory of the general linguistics (in particular, linguostatistics) while organizing the special courses, lectures and seminars, in technical means of training for a convenient placement of keys for preparation of textbooks and manuals, and compiling various dictionaries. Results of the research are practically of great value in the domain computer linguistics.

Degree of embed and economic effectivity: author's recommendations are used as practical manual on forming and developing of linguostatistics in the Uzbek linguistics. The results of research can be used also in development of the Uzbek linguistic bank (machine fund of the Uzbek language).

Sphere of usage: in the training courses in teaching subjects. "General Linguistics", "Structural, Applied and Mathematical Linguistics", "The Uzbek Linguistics", "Theory of Linguistics", "Applied Mathematics" and "Computer Linguistics".

Тадқиқотчи:

Босишга рухсат этилди 03.11.2008 й.
Бичими 60x84 1/16. Шартли 3,0 б.т. Адади 100 нусхада. Буюртма №21.
МЧЖ "MEDIANASHR" босмахонасида чоп этилди.
Манзил: Тошкент шаҳри, Ф.Мавлонов кўчаси 1а-уй.