

ДОНИШГОҲИ МИЛЛИИ ТОҶИКИСТОН

ТДУ: 004.5.7.9+519.25+811.21/.22

ТБК: 22.18 (2Т)

П - 33

Бо ҳуқуқи дастнавис



ПИРОВ САИДАЛИ МИРЗОАЛИЕВИЧ

**ТАҲИЯ ВА ТАДҚИҚИ КРИТЕРИЯҲОИ МИҚДОРӢ БАРОИ
БАҲОДИҲИИ ХУШОҲАНГИИ КАЛИМАҲОИ ТОҶИКӢ**

А В Т О Р Е Ф Е Р А Т И

**диссертатсия барои дарёфти дараҷаи илмии номзади илмҳои техники
аз рӯйи ихтисоси 05.13.11 –Таъминоти математикӣ ва барномавии
мошинҳои ҳисоббарор, комплексҳо ва шабакаҳои компютерӣ**

Душанбе, 2025

Кори илмӣ дар кафедраи технологияҳои иттилоотӣ ва иртиботии факултети механикаю математикаи Донишгоҳи миллии Тоҷикистон иҷро шудааст.

Роҳбарони илмӣ:

Усмонов З.Қ. – доктори илмҳои физикаю математика, академики АМИТ, профессор, Мирзоев С.Ҳ. – доктори илмҳои техникӣ, и.в. профессори кафедраи информатикаи Донишгоҳи миллии Тоҷикистон

Муқарризони расмӣ

Муминов Баходир Болтаевич – доктори илмҳои техники, профессор, мудири кафедраи зеҳни сунъии Донишгоҳи давлатии иқтисодии Тошкент, Ҷумҳурии Узбекистон

Ғафуров Миршафӣ Ҳамитович - номзади илмҳои техникӣ, дотсенти кафедраи технологияҳои иттилоотӣ ва ҳифзи маълумоти Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ

Муассисаи пешбар

Институти математикаи ба номи А. Ҷӯраеви АМИТ

Ҳимоя «05» сентябри соли 2025 соати «14:00» дар ҷаласаи Шурои диссертатсионии 6D.КОА-049 дар назди Донишгоҳи техникии Тоҷикистон ба номи академик М.С.Осимӣ аз рӯйи нишони: 734042, ш. Душанбе, хиёбони академикҳо Раҷабовҳо, 10 баргузор мегардад.

Бо диссертатсия дар Китобхонаи марказии Донишгоҳи техникии Тоҷикистон ба номи академик М.С.Осимӣ ва тавассути сомонаи <http://ttu.tj> шинос шудан мумкин аст.

Автореферат «__» _____ соли 2025 мувофиқи феҳристи тасдиқшуда ғиристонда шуд.

Котиби илмӣ Шурои диссертатсионии 6D.КОА-049, номзади илмҳои техникӣ, дотсент



Султонзода Ш.М.

МУҚАДДИМА

Мубрамии мавзуи таҳқиқот. Мубрамияти мавзуи таҳқиқоти пеш аз ҳама бо Қарори Ҳукумати Ҷумҳурии Тоҷикистон “Дар бораи тасдиқи барномаи истифода ва рушди технологияи иттилоотӣ дар забони тоҷикӣ”, эълон гардидани солҳои 2020-2040 “Бистсолаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илм ва маориф” ва як қатор барномаву стратегияҳои дар ин самт қабул гардида алоқамандии бевосита дорад. Таърих гувоҳ аст, ки дар ҳар давраи замон олимону муҳаққиқон ба масъалаи покизагии забон, диққати махсус додаанд ва имрӯз ҳам ин масъулияти асосиро идома додаистодаанд. Ҳифзи забон аз нуқсонҳои гуногуни зоҳирӣ, хусусан ҳамнишиниву ноҷурии овозҳои таркиби калима муҳимтарин вазифаи ҳар як фарди ҷомеа аз ҷумла эҷодкорону зиёӣён буда таъмини форамию покизагии он яке аз талаботҳои хушохонгӣ дар илми сухансанҷӣ мебошад. Ин нуқта ба тафсил дар китоби Пешвои миллат “Суханҳои ҳикматомӯзи Президент Эмомалӣ Раҳмон дар бораи забон” бараъло қайд гардидааст, ҳикмате аз онро ин ҷо овардан айнаи муддао аст: “Поку беолоиш нигоҳ доштани забон ва асолату шевоии онро ҳифз кардан қарзи фарзандии ҳар яки мову шумост. Бинобарин олимону зиёӣён мо бояд бо таҳқиқи масъалаҳои муҳимтарини забон ва таҷдиду таҳияи қомусу фарҳангномаҳо, ки барои баланд бардоштани сатҳи забондонӣ ва маърифати забаномӯзии наслҳои имрӯзу оянда заруранд, ҷиддӣ машғул шаванд” [1].

Бо пайдо гардидани аввалин барномаҳои компютерӣ, андешаи коркард, нигоҳдорӣ, таҳлил, талаффуз ва тарҷумаи матнҳо низ ба вучуд омад. Барномаҳои гуногун барои коркарди иттилооти матнӣ таҳия гардидаанд, ки муҳимтарин масъалаҳои илми забоншиносиро ҳал мекунанд. Яке аз масъалаҳои мазкур ин хушохонгӣ ё хоноии матнҳо ва баҳодихии он тавассути компютер мебошад, ки он барои қисме аз забонҳои дунё ҳали худро пайдо намудааст. Масалан, дар дастаи барномаҳои оффисии Office, ки маҳсули ширкати Microsoft мебошад, афзоре мавҷуд аст, ки омори муносибхонии матнро пешниҳод мекунад ва ғайр аз баҳодихии хониши осон ва ҳисоб

кардани миқдори калимаҳои душворталаффуз, инчунин қобилияти арзёбӣ кардани хушоҳангиро низ дорад.

Калима, ибора ва ё матн хушоҳанг номида мешавад, ки агар дар таркиби он ҳамсадоҳои нодир, якмахрача ё махрачи наздик набошанд. Дар таркиби калимаҳои хушоҳанг садонокҳо нисбат ба ҳамсадоҳо бештар буда, то се ҳиҷои кушодро дошта метавонанд. Баҳри ҳифз намудани покизагии хушоҳангии забони тоҷикӣ аз душворталаффузӣ, танофурот ва чигилӣ зарур аст, ки аз тамоми роҳу воситаҳо, хусусан аз технологияҳои иттилоотии муосир, истифода карда меёра, амсилаи математикӣ ва низомҳои компютери маҳсус таҳия карда шаванд. Таҳқиқоти мазкур асосан ба таҳияи асосҳои низоми шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ равона гардидааст, ки асоси онро ӯ-таснифгар ташкил медиҳад.

Гарчанде, ки меёраҳои хушоҳангӣ дар баъзе забонҳо ба расмӣ дароварда нашудаанд, вале онҳо мавҷуд мебошанд. Масалан, муҳимтарин меёраҳои хушоҳангии матн: сохтори ҳиҷо, таносуби овозҳои пасту баланд, сахми ҳамсадоҳои сонорӣ, овозҳои хуштакию шавшувдор ва якҷояшавии ҳамсадоҳо ба ҳисоб мераванд. Масъалаи хушоҳангии матн масъалаи хеле мураккаб буда, ҳатто одамони як минтақа метавонанд, ки дар бораи зебогии нутқ фаҳмиши гуногун дошта бошанд. Ин ҷо омилҳои зиёдеро ба ҳисоб гирифтани зарур аст, ки дар асоси онҳо пойдевори фаҳмиши ҳар як инсон дар алоҳидагӣ сохта мешавад. Ғайр аз ин, як қатор хусусиятҳои низ мавҷуд аст, ки онҳо сифати иттилооти ин ё он сарчашмаро ҳалалдор меkunанд (масалан, дар матн метавонад ҳуруф, дар китоб сифати қоғаз, дар нутқ бошад овоз ва оҳанг).

Дарачаи таҳқиқи мавзӯи илмӣ. Сарчашмаҳо нишон медиҳанд, ки аввалин шуда мутафаккирони Юнони қадим суҳангӯию суҳансанҷиро шуруъ намуда, онро мавриди омӯзишу баррасӣ қарор додаанд. Асари “Риторика” (Суҳанварӣ) – и Арасту далели равшани ин гуфтаҳо мебошад. Дар хусуси масъалаи покизагии забон мутафаккирони машриқзамин аз қабилҳои Абӯалӣ ибни Сино, Шамс Қайси Розӣ, Асадии Тӯсӣ, Шамси Гургонӣ, Т. Зеҳнӣ, Х. Шарифов, М. Шукуров, Б. Камолиддинов ва дигарон қорҳои зиёдеро анҷом

додаанд, ки муҳимтарин сарчашмаҳои омӯзиши паҳлуҳои гуногуни масъалаи мазкур ба ҳисоб мераванд.

Аз оғози асри гузашта то имрӯз хусусиятҳои гуногуни фонетикаи забони адабии ҳозираи тоҷикиро олимони машҳури ин соҳа И.И. Зарубин, Л. Бузургзода, В. С.Соколова, В.С. Расторгуев, Ш. Бобомуродов, Т.Н. Хаскашев, М. Файзов, А.Неъматов, Ш.Кабиров ва шогирдонашон мавриди омӯзишу баррасӣ қарор додаанд. Саҳми шодравон профессор Т.Н. Хаскашев ва М. Файзов дар омӯзишу инкишофи фонетикаи забони адабии муосири тоҷик хеле зиёд мебошад. Ин муҳаққиқони ватанӣ таҳлилу муқоисаҳои илмӣ, назариявӣ ва амалии арзишмандеро дар илми фонетикаи тоҷик анҷом дода, хусусиятҳои тавлидӣ (физиологӣ) ва шунавоии (акустикии) овозҳои забони тоҷикиро бо истифодаи асбобҳои (рентгенограмма, политограмма) махсуси дар соҳаи тиб ва ғайра истифодашаванда ошкор намудаанд.

Ба омӯзиши масъалаи мазкур аз нигоҳи илми риёзӣ ва технологияҳои иттилоотӣ дар хориҷи кишвар олимони зиёде ба мисли Rudolf Flesch, Gunning R., Kincaid J., Spache G., Fry E., Hedman S., McLaughlin, G., В. Франчина, Р. Вакка, Оборнева И. В., Иванов К.В., Фомин А.Ю. ва дигарон машғул гардида, амсилаҳои математикии баҳодихӣ ва таҳлили хушоҳангии матнҳои англисӣ, итолявӣ, русӣ ва забонҳои дигарро аз рӯи параметрҳои гуногун омода намуда, натиҷаҳои назаррасеро ба даст овардаанд, ки дар калонтарин ширкатҳои ҷаҳонӣ, аз қабيلي Microsoft ва дигар ширкатҳо, пайкараҳои миллии забонҳо, сомонаҳои ҷуғрофӣ ва байналмилалӣ истифода мешаванд.

Дар Ҷумҳурии Тоҷикистон дар самти коркарди забони табиӣ, махсусан забони тоҷикӣ муҳаққиқон Усмонов З.Ҷ., Исмоилов М.А., Тошхучаев Х.А., Умаров М.А., Зарипов С.А., Шарифов Ш., Гращенко Л.А., Худойбердиев Х.А., Солиев О.М., Довудов Г.М., Қосимов А.А., Қаюмов М.М. ва дигарон якҷоя бо шогирдонашон таҳқиқот бурда истодаанд.

Таҳияи низомҳои шинохти худкори хушоҳангии калимашаклҳои имконияти сохтани фарҳанги луғавии калимашаклҳои хушоҳангро медиҳад, ки он дар оянда метавонад қисми зарурии пайкараи матнҳои забони тоҷикӣ

гардад. Муваффақиятҳо дар самти таҳияи шинохти худкори хушоҳангии калимашаклҳо замина ба рушди босуръати низомҳои худкори баҳодихии хушоҳангӣ гузошта, имкон медиҳад, ки хушоҳангиву душворталаффузии ҷумлаҳо, матнҳо, асарҳо, китобҳои дарсӣ, матни имтиҳонҳои хонандагону донишҷӯён ва эҷодиёти ҳаргуна муаллифон арзёбӣ гарданд.

Робитаи таҳқиқот бо барномаҳо (лоиҳаҳо) ва мавзӯҳои илмӣ. Кори диссертатсионӣ дар доираи нақшаи корҳои илмӣ - таҳқиқотии кафедраи технологияи иттилоотӣ ва иртиботии факултети механикаю математикаи Донишгоҳи миллии Тоҷикистон тибқи амалишавии нақшаи илми-таҳқиқотии мавзӯҳои бучавӣ, рақами қайди давлатӣ № 0122TJ14117 (01.01.2021-31.12.2025) «Татбиқи методҳои риёзӣ ва технологияҳои нави иттилоотӣ дар соҳаи забоншиносӣ ва омӯзиши методҳои рамзгузори алифбой» ба қайд гирифта шудааст ва лоиҳаи тадқиқоти бунёдии рақами қайди давлатӣ № 0116TJ00661 (01.01.2016-31.12.2021) «Татбиқи методҳои моделсозии математикӣ ва технологияҳои нави иттилоотӣ дар соҳаи маориф ва забоншиносӣ», иҷро гардидааст.

ТАВСИФИ УМУМИИ ТАҲҚИҚОТ

Мақсади таҳқиқот. Таҳияи асосҳои низомҳои худкори шинохти хушоҳангии калимашаклҳои забони тоҷикӣ ва амалӣ намудани он дар намуди маҷмааи барномаи компютерӣ.

Вазифаҳои таҳқиқот. Барои ба мақсад расидан масъалаҳои зерин ҳал карда мешаванд:

- таҳияи маҷмуаи матнии калонҳаҷм барои омӯзиш ва гузаронидани озмоишҳои гуногуни илмӣ аз ҷумла муайянкунии хушоҳангии калимашаклҳо;
- таҳқиқи иттилоъдиҳии аломатҳои, ки барои муайянкунии хушоҳангии калимашаклҳо таъйин шудаанд;
- тафтиши самаранокии усулҳои математикӣ барои муайянкунии

хушоҳангии калимашаклҳо;

- сохтани лоиҳаи маҷмаи барномаи компютерӣ барои ба таври худкор муайян кардани хушоҳангии калимашаклҳои тоҷикӣ ва татбиқи он.

Объекти таҳқиқот – амсилаҳои математику компютери таҳлил ва баҳодиҳии хушоҳангии маҷмуи калимашаклҳои ҷопии забони тоҷикӣ.

Мавзӯи таҳқиқот – шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ дар асоси аломатҳо ва басомади N -граммаҳои он.

Усулҳои таҳқиқот. Барои ҳалли мақсаду вазифаҳои дар тадқиқот гузошташуда аз методҳои фонетикӣ, таҳлили омӯрӣ, ҳисобҳои таҷрибавӣ, назарияи маҷмуӣҳо, таҳлили системавӣ, амсиласозии математикӣ ва компютерӣ, барномасозии ба объект нигаронидашуда барои таҳияи воситаҳои барномавӣ ва коркарди манбаи додаҳои релятсионӣ истифода шудааст.

Асосҳои назариявӣ таҳқиқот. Асосҳои назариявӣ диссертатсияро сарчашмаҳои илмӣ ниёгон, натиҷаҳои илмӣ олимони давраи замони ғуноғун, ки ба мавзӯи диссертатсия иртибот доранд, ташкил медиҳанд.

Навгони илмӣ таҳқиқот. Натиҷаҳои асосии диссертатсия нава буда, ҷунин ифода мегарданд:

- иттилоъдиҳии аломатҳое, ки барои шинохт ва муайянкунии хушоҳангии калимашаклҳо таъйин шудаанд, таҳқиқ карда шуд;

- самаранокии истифодаи N -граммаҳои ҳуруфӣ ($N=1,2,3$) ба сифати аломатҳои муайянкунандаи хушоҳангии калимашаклҳо, муқаррар карда шуд;

- самаранокии таснифгари Усмонов З.Ҷ. муқаррар карда шуд, ки қобилияти ба ҳаётии то 96% муайян кардани хушоҳангии маҷмуи калимашаклҳоро дорад;

- бори аввал дар Тоҷикистон маҷмаи барномаи компютери замонавӣ барои шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ, сохта шуд.

Нуктаҳои ба ҳимоя пешниҳодшаванда. Нуктаҳои зерин ба ҳимоя пешниҳод карда мешаванд:

1. Аломатҳое, ки барои шинохтан ва муайян кардани хушоҳангии калимашаклҳо таъйин шудаанд;

2. Тасвири миқдории калимашаклҳои тоҷикӣ, ки хушоҳангиро дар намуди тақсимои басомади N-граммаҳои ҳуруфӣ ($N=1,2,3$) тавсиф мекунанд;
3. Татбиқи самаранокии таснифгари Усмонов З.Қ., ҳамчун методи нав ва мувофиқ барои муайян кардани хушоҳангии калимашаклҳои тоҷикӣ;
4. Маҷмаи барномаи компютерии замонавӣ барои шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ.

Аҳамияти назариявӣ таҳқиқот. Аҳамияти назариявӣ диссертатсия дар он зухур меёбад, ки дар он бо усули таҷрибавӣ, усули нави таснифоти бузургҳои тасодуфӣ фосиладор мавриди санҷиш қарор ёфт ва самаранокии табиқи он бо мақсади шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ, муқаррар карда шуд.

Аҳамияти амалӣ таҳқиқот. Аҳамияти амалӣ қор аз он иборат аст, ки вай барои табиқи амсилаҳои риёзӣ иттилоотӣ ва маҷмаи барномаи компютерии дар он сохташуда дар соҳаҳои илму маориф, фарҳанг, забоншиносӣ, адабиётшиносӣ ва низомҳои компютерии қорқарди матн барои шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ пешбинӣ шудааст.

Саҳеҳият ва асоснокии натиҷаҳои бадастомада, бо як силсила озмоишҳои ҳисоббарорӣ тасдиқ мешаванд, ки дар онҳо хушоҳангии калимашаклҳои гуногун бо саҳеҳияти қифоя баланд муайян шудаанд.

Мутобиқати диссертатсия ба шиносномаи ихтисоси илмӣ.

Мундариҷаи таҳқиқоти диссертатсия ба бандҳои 1, 4 ва 7-и шиносномаи ихтисоси илмӣ барои дарёфти дараҷаи илмӣ номзади илмҳои техникӣ аз рӯи ихтисоси 05.13.11 – «Таъминоти математикӣ ва барномавии мошинҳои ҳисоббарор, комплексҳо ва шабакаҳои компютерӣ», мувофиқат мекунад:

Б.1. Амсилаҳо, усулҳо ва алгоритмҳои банақшаگیرӣ ва таҳлили барнома ва низомии барномавӣ, инчунин тағйирдиҳии эквивалентӣ, верификатсия ва тестирунии онҳо;

Б.4. Низомии идоракунии манбаи маълумот ва дониш;

Б.7.Интерфейсҳои инсон-мошин, амсилаҳо, усулҳо, алгоритмҳо ва воситаҳои барномавии мошинии графикӣ, визуализатсия, коркарди тасвир, низоми ҳақиқати виртуалӣ, муоширати мултимедӣ.

Саҳми шахсии доктарабон дарачаи илмӣ дар таҳқиқот. Гузориши масъала ва интихоби методҳои таҳқиқот ба роҳбарони илмӣ тааллуқ дошта, натиҷаҳои асосии қори диссертатсионӣ аз тарафи муаллиф мустақилона ба даст оварда шудаанд.

Тасвиб ва амалисозӣ. Натиҷаҳои асосии қор дар конференсияҳо ва семинарҳои зерин пешниҳод ва муҳокима гардидаанд:

- семинарҳои илмӣ-тадқиқотӣ дар Донишгоҳи миллии Тоҷикистон, Институти математикаи АМИТ, Донишгоҳи техникии Тоҷикистон ба номи М.С. Осимӣ, Донишгоҳи технологии Тоҷикистон ва Донишгоҳи славянии Тоҷикистону Руссия дар солҳои 2014-2023;

- конференсияи илмӣ-амалии байналмилалии "Перспективы развития науки и образования", соли 2016., Душанбе;

- конференсияи илмӣ-амалии байналмилалии “Нақши технологияҳои иттилоотӣ коммуникатсионӣ дар рушди инноватсионии иқтисодии Ҷумҳурии Тоҷикистон”, соли 2017, шаҳри Душанбе;

- конференсияи байналмилалии илмӣ “Современные проблемы математики и их приложения”, 14-15 июни соли 2017, шаҳри Душанбе, шаҳри Кӯлоб;

- Конференсияи илмӣ-амалии байналхалқии «XII Ломоносовские чтения», посвященной Дню таджикской науки и 30-летию установления дипломатических отношений между Республикой Таджикистан и Российской Федерацией (29-30 апреля 2022). Часть I. Естественные науки. – Душанбе, 2022;

- Материалы международной научно-практической конференции “Современные проблемы математики и её приложения”, посвященной 20-летию развития естественных, точных и математических наук 2020-2040 годы . -Душанбе -2022;

- Маводҳои конференсияи байналмилалӣ илмӣ-амалӣ дар мавзуи «Таҳлили компютери масъалаҳои илм ва технология» бахшида ба «Солҳои 2020-2040 эълон гардидани 20-солаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илму маориф» ва «75-солагии Донишгоҳи миллии Тоҷикистон». -Душанбе -2023.

Интишорот аз рӯйи мавзуи диссертатсия. Аз рӯйи мавзуи диссертатсионӣ 13 мақола нашр шудаанд, [1-М - 13-М]. Аз онҳо 6 мақолаи илмӣ, дар нашрияҳои аз тарафи Комиссияи олии аттестатсионии назди Президенти Ҷумҳурии Тоҷикистон тавсияшуда, нашр гардидаанд.

Сохтор ва ҳаҷми диссертатсия. Рисолаи илмӣ аз мундариҷа, се боб, ҳулоса ва рӯйхати адабиёт, ки аз 137 номгӯй иборат аст, таркиб ёфтааст. Муҳтавои диссертатсия дар 143 саҳифа дарҷ гардида, 13 ҷадвал ва 59 расмро дар бар мегирад.

МУҲТАВОИ АСОСИИ ТАДҚИҚОТ

Дар **муқаддима** аҳамияти мавзуи интихобгардида асоснок карда шуда, ҳадаф, вазифаҳо, объект ва мавзуи тадқиқот муайян шудаанд, навлонӣ ошкор, аҳамияти назариявӣ ва амалии кор муаррифӣ шудааст, инчунин муқаррароти ба ҳимоя пешниҳодшуда, усулҳои таҳқиқ, асос ва маводи омӯзиш тавсиф дода шуда, дар бораи тасвиби тадқиқот ва сохтори он маълумот оварда шудааст.

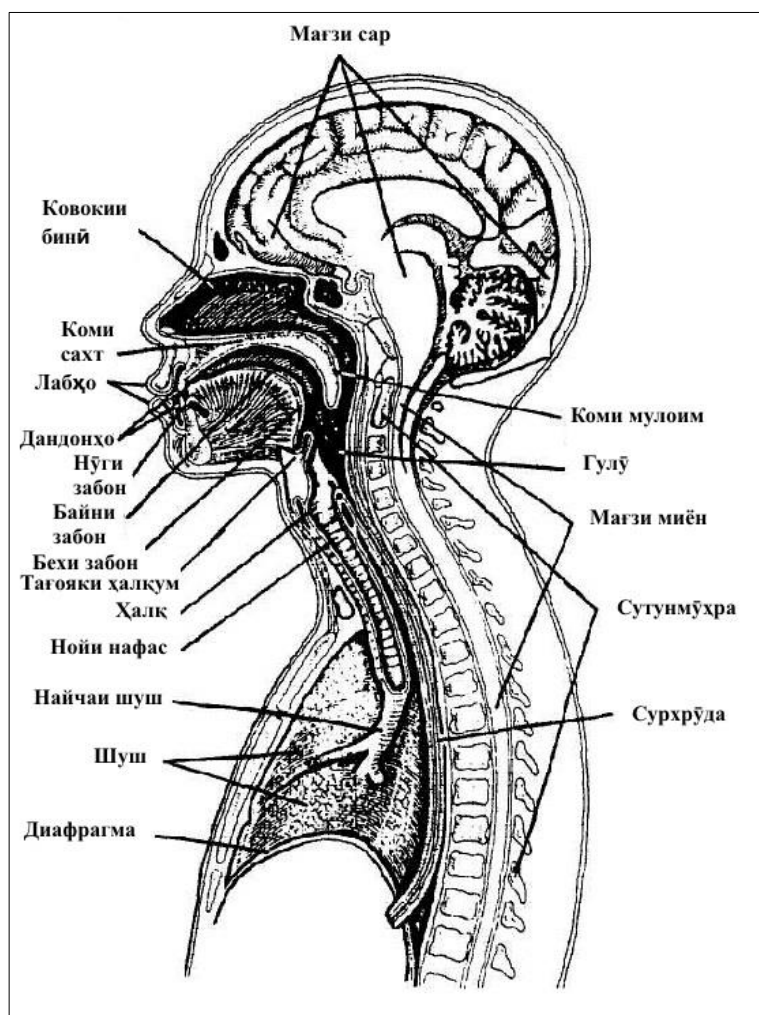
Дар **боби 1** масъалаи ҳушоҳангии матнҳо, тарзҳои ҳалли он, ҷанбаҳои пайдоиши овозҳои нутқ ва хусусиятҳои физикию тавлидии онҳо таҳлил карда мешаванд. Дар **зербоби 1.1** андешаҳои муҳаққиқони давраи замони гуногун аз соҳаҳои филология, математика ва технологияҳои иттилоотӣ дар бораи масъалаи ҳушоҳангӣ ва душворталаффузии калимаҳо мавриди таҳқиқ қарор гирифтаанд.

Дар **зербоби 1.2** ҷанбаҳои пайдоиши овозҳои нутқ баррасӣ шудаанд. Сохти овозии забонҳои муоширати байни одамон табиати бисёр мураккаб дошта, аз ин рӯ зарурати омӯзиши паҳлуҳои гуногуни он ба миён меояд. Нутқи

мо, пеш аз ҳама, ҳодисаи овозӣ буда, табиати физикӣ дорад. Ҳар як ҳодисаи овозӣ дар вақту замони муайян бо қувваю зудии ба худ хос ба вуқӯъ меояд.

“Дар заминаи физику акустикӣ табиати физикуи овозҳои нутқ ва аломатҳои онҳо: дарозӣ, дараҷаи паҳннокӣю қувватнокӣ, сохти формантию спектрӣ мавриди таҳлилу таҳқиқ қарор мегиранд. Ин аломатҳои физикуи овозҳои нутқ дар шароит ва ҳолатҳои гуногуни фонетикӣ таҳқиқу муқоиса ва дараҷаи молик будани онҳо ба ин хусусиятҳо нисбат ба овозҳои дигар таъмин карда мешавад. Таснифи акустикӣю овозҳо низ дар асоси ҳамин аломатҳо сурат мегирад”.

Овозҳое, ки мо мешунавем, манбаъ ва сабаби пайдоиш доранд. Манбаъ ва сабаби пайдо гардидани овозҳои нутқ ин дастгоҳи гуфтори инсон ва қори узвҳои он аст (расми 1.2.1).



Расми 1.2.1 - Дастгоҳи гуфтор

Қисми асосии халқ, ки дар овозсозӣ (садосозӣ) саҳми асосӣ дорад садопардаҳо мебошанд. Садопардаҳо қодиранд, ки дар як сония тахминан аз 40 то 1700 маротиба ларзанд, вале дар нутқи маъмулӣ лапиши онҳо ба 85-200 (дар нутқи мардон) ё 160-340 (дар нутқи занон) ларза дар сония баробар аст. Садопардаҳои мардон ғафс ва аз занон тунук мебошанд, ки аз ин хотир миқдори ларзишҳои садои мардон паст ва аз занон баланд мебошад.

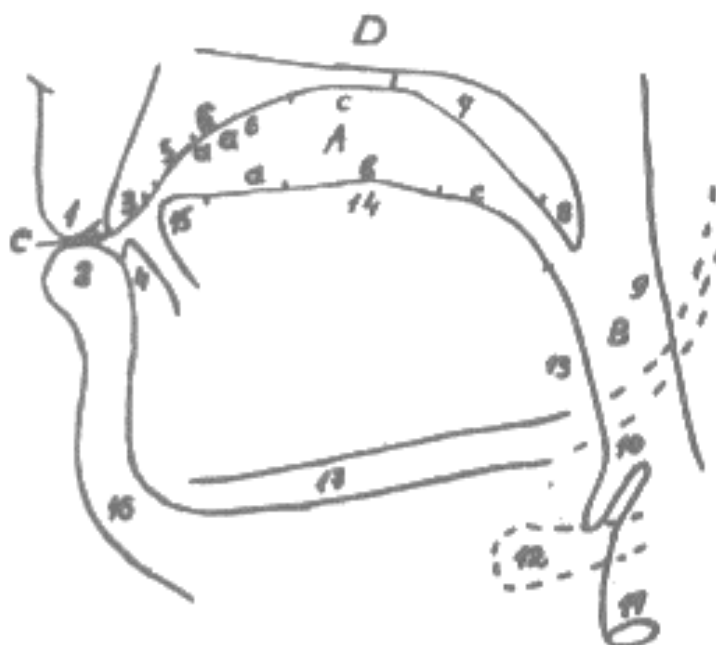
Ҳар як забон дорои низоми овозҳои ба худ хос аст, ки онҳо ба воситаи узвҳои гуногун сохта мешаванд. Узвҳои овозсозии ғайрифоҳи забони тоҷикӣ: садопардаҳо, деворҳои халқ, коми мулоим бо забонча, забон ва лабҳо мебошанд, ки дар ташаккули овозҳо саҳми бештар доранд. Масалан, аз 30 фонема (овоз)-и забони тоҷикӣ 20 фонема (66,7%) дар қисми пеши забон ва танҳо 10 фонемаи дигар: садонокҳои у, ў, о ва ҳамсадоҳои г, к, х, ғ, қ, ъ, ҳ (33,3%) дар қисми қафои забон сохта мешаванд. Ба узвҳои овозсозии ғайрифоҳӣ: коми саҳт, дандонҳо, милкҳои дандонҳо ва ковокии бинӣ дохил мешаванд.

Гарчанде ки хусусиятҳои дастгоҳи овозсозии забони тоҷикӣ ҳанӯз ҳам таҳқиқ нашуда бошад ҳам оиди баъзе аз онҳо маълумоти муқаддимагӣ мавҷуд аст. Яке аз ин гуна хусусиятҳои, ки ба дастгоҳи овозсозии забони тоҷикӣ тааллуқ дорад, хусусияти мусиқагӣ - ҳушовозию ҳушохангӣ мебошад. Ҳушохангӣ ин ҳамнишинии комили овозҳои мебошад, ки барои гӯянда осонӣ ба шунаванда форами меорад. Талаботҳои ҳушохангӣ бояд мутобиқи хусусиятҳои фонетикии забон мушаххас бошанд. Ҳушохангӣ ҳамеша ба гуногунрангии фонетикаи як забони миллии вобастагӣ дорад.

Дар **зербоби 1.3** хусусиятҳои физикии овозҳо, хусусан овозҳои забони тоҷикӣ мавриди омӯзишу баррасӣ қарор гирифтаанд.

Дар **зербоби 1.4** суҳан дар бораи хусусиятҳои тавлидии овозҳои забони тоҷикӣ меравад. Алифбои ҳозираи забони адабии тоҷикӣ аз 35 ҳарф ва 30 овоз (фонема) иборат аст, ки онҳоро ба ду гурӯҳи асосӣ: 1) садонокҳо ва 2) ҳамсадоҳо ҷудо мекунанд. Дар талаффузи садонокҳо монеа дида намешавад.

Ғайр аз ин садонокҳо бо дигар хусусиятҳояшон аз ҳамсадоҳо фарқ мекунанд. Чун дар талаффузи садонокҳо монеа дида намешавад, гуфтан мумкин аст, ки онҳо овозҳои форам ҳастанд. Ҳамсадоҳо нисбатан бештар буда, нақши зиёдро дар ташкили калима ва матн доранд. Ҳангоми талаффузи ҳамсадоҳо бошад, онҳо ба монеаҳои гуногун дучор мешаванд, ки ин то як андоза роҳи чараёни овозро гирифта вазниниро дар талаффуз пеш меорад. Масъалаи ҳалалдор шудани хушоҳангии калимашаклҳо низ аз ҳамнишиниву теъдоди ҳамсадоҳо вобастагӣ дорад. Барои равшантар шудани гуфтаҳоямон нақшаи узвҳои нутқро меорем, ки тавассути асбоби рентгенограмма гирифта шудааст (расми 1.4.1).



Расми 1.4.1 – Инъикоси узвҳои нутқ дар рентгенограмма

Дар нақшаи рентгении расми 1.4.1 бо рақамҳо узвҳои талаффуз ва бо ҳарфҳо ковокиҳои дастгоҳи нутқ ишора карда шудааст. Ковокии даҳон – А, ковокии ҳалқ – В, ковокии даромадгоҳи даҳон – С, ковокии бинӣ – D. Бо рақамҳои 1 – лаби боло; 2 – лаби поён; 3 – дандонҳои боло; 4 – дандонҳои поён; 5 – садопардаҳо; 6 – коми саҳт: а) қисми пеш, в) қисми мобайн, с) қисми кафо; 7 – коми нарма; 8 – забонча; 9 – девораи ҳалқ; 10 – тағояки болои ҳалқ; 11 – ҳалқ; 12 – устухони зери забон; 13 – решаи асоси забон; 14 – пушти забон.

а) пеши забон, в) байни забон, с) қафои забон; 15 – нӯги забон; 16 – манаҳ; 17 – чоғи поён ишора гардидаанд.

Дар **зербоби 1.5** боби якуми рисола хулосаҳо оид ба боби якум оварда шудаанд.

Дар **боби 2** ташаккули асосҳои андозагирии хушоҳангӣ, татбиқи амсилаи математикӣ ва компютерӣ бо истифодаи элементҳои униграмма, биграмма ва триграммаҳо дида баромада шудааст.

Дар **зербоби 2.1** ташаккули асосҳои андозагирии хушоҳангӣ дида баромада шудааст. “Хушоҳангӣ (эвфония) мафҳуми сифатӣ мебошад, зеро ҳар фард ба сазои калима, ҷумла, шеър, матн ва гуфтор ба таври худ баҳо медиҳад. Дар баробари ин, мафҳуми баррасишаванда омӯрӣ аст, зеро маҷмуи шахсони алоҳида нисбати як объект дар бисёр мавридҳо андешаи «умумӣ»-ро ифода мекунанд. Ин далел асоси рушди эвфонометрия ҳамчун фанни нави илмӣ, ки ба омӯзиши қолабҳои хушоҳангӣ дар нутқи хаттӣ ва шифохӣ бо истифода аз усулҳои риёзӣ ва омӯрӣ нигаронида шудааст, ба ҳисоб меравад. Аҳамияти амалии чунин тадқиқотҳо дар таъсири бешубҳаи сифати овозии нутқ ба натиҷаҳои дарки иттилоот мебошад. Ташаккули асосҳои эвфонометрия аз сохтан ё мурочиат кардан ба маҷмуаи матнии аллакай мавҷудбудаи ягон забони табиӣ (беҳтараш ба пайкараи забонӣ) оғоз меёбад. Маҷмуа бояд ба қадри кофӣ намоён буда фарогири адабиёти бадеӣ, мақолаҳои рӯзномаю маҷаллаҳо, матнҳои илмию оммавӣ аз соҳаҳои гуногуни дониш, асарҳои ҷамъиятию сиёсӣ, намоишнома ва ғайраҳо бошад. Мувофиқи матлаб аст, ки ҳаҷми маҷмуа аз даҳҳо миллион калима иборат бошад” [1 – М].

Дар асоси 20 асари адибони форсу тоҷик аз давраи замони мухталиф ва осори дорони жанрҳои гуногун маҷмуаи матниеро тартиб додем, ки аз 6 811 саҳифаи электронӣ иборат буда 2 206 587 калимаро дар бар мегирад. Натиҷаи омории ин пажӯҳиш дар ҷадвали 2.1.1 оварда шудааст.

Ҷадвали 2.1.1 - Омори маҷмуаи матнии таҳқиқшаванда

№	Номи адиб	Номи асар	Миқдори саҳифа	Миқдори калима
1	Абдуррамони Ҷомӣ	Баҳористон	139	31 934
2	Саъдии Шерозӣ	Гулистон	150	39 589
3	Носири Хисрав	Сафарнома	100	34 523
4	Ибни Муқаффаъ	Калила ва Димна	120	39 647
5	Садриддин Айний	Марги судхӯр	220	57 926
6	Ҳусайн Воизи Кошифӣ	Футувватнома	336	93 874
7	Мухаммадиев Ф.	Шаби савум	288	94 952
8	Меҳмон Бахтӣ	Тилисми Сайхун	45	12 020
9	Сорбон	Романи Туғрал	190	57 876
10	Мухаммадҷони Шакурии Бухорӣ	Садри Бухоро	188	112 709
11	Мухаммадҷони Шакурии Бухорӣ	Хуросон аст ин ҷо	361	92 104
12	Ҷалолуддини Балхӣ	800 ғазали дилошӯб	287	83 215
13	Эмомалӣ Раҳмонов	Нигоҳе ба таърих ва тамаддуни ориёӣ	371	144 387
14	Ғафуров Б.Ғ.	Тоҷикон	870	305 673
15	Лоик Шералӣ	Куллиёт	830	210 713
16	Абулқосим Фирдавсӣ	Шоҳнома. Ҷилди 1	277	57 694
17	Худойназар Асозода	Достони зиндагӣ (китоби 4-ум)	321	145 205
18	Худойназар Асозода	Достони зиндагӣ (китоби 5-ум)	480	223 087
19	Абулфазли Байҳақӣ	Таърихи Байҳақӣ	756	235 890
20	Маҷлиси Олии ҚТ	Кодекси андоз	482	133 569
			6811	2 206 587

Ба се синф тақсим кардани калимашаклҳо дар асоси пурсиши афкори умум ба воситаи пурсишнома (анкета) сурат гирифт. Мусоҳибон ба ростӣ ҳар як калимашакли аз рӯйхати $\bar{L}$ яке аз се рақами: -1, +1 ва 0 гузоридаанд, ки онҳо аломатҳои сифатии мувофиқи овозҳои «душвор», «форам» ва «миёна»-ро кодгузорӣ мекунанд. Ба мутахассисон (соҳибони забон) калимашаклҳои интихобшударо пешкаш намудем, онҳо ба ҳар як калима яке аз се рақами: -1, 0 ё +1 - ро гузоштанд, ки мувофиқан сифати аломатҳои талаффузи “душвор”, “миёна” ва “форам” - ро кодгузорӣ мекунанд. Пурсишнома бо ҳаҷми шаш саҳифаи ҷопӣ дар шакли ҷадвал пешкаши мутахассисон гардид.

Мутахассис калимашаклҳои дар анкетабударо аз назар мегузаронад, агар талаффузи он душвор бошад, онгоҳ ба ростӣ он рақами -1, агар талаффузи калимашакл на душвору на форам, миёна бошад, онгоҳ ба ростӣ он рақами 0

ва агар талаффузи калима форам бошад, онгоҳ дар паҳлӯи он рақами +1 ро менависад.

Дар асоси хусусиятҳои гуногуни овозҳои забони тоҷикӣ, қонуниятҳои ҳамнишинии онҳо, пурсиши соҳибони забон ва коркарди компютери калимашаклҳо 16 аломатро ошкор намудем, ки тавсифкунандаи сифати гуногуни овозии калимашаклҳои забони адабии тоҷик мебошанд. Аломатҳое, ки барои тавсифи миқдорию сифатии хушоҳангии калимашаклҳо интиҳоб гардидаанд дар ҷадвали 2.1.2 бо қиматҳои имконпазир оварда шудаанд.

Ҷадвали 2.1.2. - Аломатҳои тавсифкунандаи хусусиятҳои миқдорию сифатии калимашаклҳо

№	Аломати калимашаклҳо	Қиматҳои имконпазир	Табододи мавқеи аломат
1	Дарозӣ (миқдори ҳарфҳо)	Аз 1 то 30	2
2	Шумораи садонокҳо	аз 1 то 15	2
3	Шумораи ҳамсадоҳо	аз 1 то 15	2
4	Ҳамсадои Ғ	1-ҳаст, 0-нест	1
5	Ҳамсадои Җ	1-ҳаст, 0-нест	1
6	Ҳамсадои Қ	1-ҳаст, 0-нест	1
7	Ҳамсадои Ҳ	1-ҳаст, 0-нест	1
8	Ҳамсадои Ҷ	1-ҳаст, 0-нест	1
9	Ҳамсадои Ъ	1-ҳаст, 0-нест	1
10	Биграммаи НГ	1-ҳаст, 0-нест	1
11	Биграммаи НД	1-ҳаст, 0-нест	1
12	Биграммаи РД	1-ҳаст, 0-нест	1
13	Биграммаи СТ	1-ҳаст, 0-нест	1
14	Биграммаи ФТ	1-ҳаст, 0-нест	1
15	Биграммаи ХТ	1-ҳаст, 0-нест	1
16	Биграммаи ШТ	1-ҳаст, 0-нест	1

Дар зербоби 2.2 амсилаи шинохти хушоҳангии калимашаклҳои тоҷикӣ бо истифодаи униграммҳо дида баромада шудааст. Дар ин фасл ва фаслҳои ояндаи ин боб равиши системавӣ истифода бурда мешавад, ки соли 2015 барои аз сифр таҳия намудани асосҳои эвфонетрияи забонҳои табиӣ, аз он ҷумла забони тоҷикӣ пешниҳод шуда буд ва онро барои забонҳои табиӣ гуногун ҳам татбиқ кардан мумкин аст. Асоси ин равишро γ-таснифгари бузургҳои тасодуфии фосилдор (дискретӣ) ташкил медиҳад, ки он аз тарафи академики АМИТ Усмонов З.Ҷ. таҳия ва коркард гардидааст. Таснифгари номбурдари барои ҳалли масъалаҳои муҳталифи илми забоншиносӣ ва ошкор

намудани хусусиятҳои воҳидҳои матнии гуногун, аз ҷумла муайян кардани сифати талаффузи калимашаклҳо, истифода намудан мумкин аст. Дар ин қисмати кор таснифгари болозикрро барои муайян кардани сифати овозии калимашаклҳо истифода мекунем, ки он сегонаи математикии аз се қисм иборат бударо, пешкаш мекунад:

- симои рақамӣ (CP), яъне тавсифи миқдории калимашаклҳо;
- формулаҳои масофаи байни CP калимашаклҳо;
- алгоритми қабули қарорҳо дар асоси тимсолҳо.

Таснифгари мазкурро барои шинохти худкори сифати калимашаклҳои маҷмуи муайян: асари назмӣ ё насрӣ, ки пешакӣ аз пайкараи матнӣ интиҳоб гардидааст, истифода мебарем. Пайкараи матнӣ ин маҷмуи матнҳои дорои жанрҳои гуногун, мебошад, ки дар шакли электронӣ нигоҳдорӣ шудааст. Пайкараи матнро барои гузаронидани пажӯҳишҳои гуногуни соҳаи забоншиносӣ, таърих ва соҳаҳои дигар истифода мекунам. Аз пайкараи матнҳои тоҷикӣ маҷмуи баъзе калимашаклҳоро интиҳоб мекунем. Чунин маҷмуъ калимашаклҳои нисбатан серистеъмоли талаффузашон “душвор”, “миёна” ва “хушоҳанг” – ро дар бар мегирад.

Маҷмуи калимашаклҳои тоҷикиро бо ҳарфи G ишора карда онро аз рӯи сифати талаффузашон ба се гурӯҳ ҷудо мекунем: G_1 – душворталаффуз, G_2 – хушоҳанг ва G_3 – калимашаклҳои миёна ё бетараф. Аз ҳар як гурӯҳ 1000-1500 калимашаклро интиҳоб мекунем. Гурӯҳи G_1 -ро бо ягон тарзе ба 3 зергурӯҳи – G_{11} , G_{12} , G_{13} – бо андозаҳои тахминан баробари шумораи калимашаклҳо, ҷудо мекунем. Гурӯҳи G_2 -ро низ бо ягон тарзе ба 3 зергурӯҳи G_{21} , G_{22} , G_{23} – бо андозаҳои тахминан баробари шумораи калимашаклҳо, ҷудо мекунем. Зергурӯҳҳои G_{1i} ва G_{2j} -ро ҳамчун зергурӯҳҳои калимашаклҳо (матнҳо) дида мебароем.

“Дар қадами аввал ба сифати унсурҳои тавсифкунандаи инъикоси миқдории зергурӯҳҳои калимашаклҳо алифбои ягонаи тоҷикиро истифода мекунем, ки он аз 35 ҳарф (ё униграмма): *а, б, в, ..., э, ю, я* иборат аст.

Барои ҳар яке аз 6 зергурӯҳҳои G_{1i} ва G_{2j} симои рақамии ягонаеро дар намуди ҷадвали қонуни тақсимот, дохил мекунем

$$\begin{array}{lcl} \bar{N} & : & 1, \quad 2, \quad \dots, \quad m \\ P & : & p_1, \quad p_2, \quad \dots, \quad p_m, \end{array} \quad (2.2.1)$$

дар сатри аввал рӯйхати $m = 35$ ҳарфи тоҷикӣ бо тартиби алифбо рақамгузоришуда ва дар сатри дуюм бошад басомадҳои онҳо барои гурӯҳи мушаххас оварда шудааст, зимнан эҳтимолияти онҳо низ ба қимати ифодаи зерин баробар аст:

$$\sum_{k=1}^m p_k = 1.$$

Ба ҳар яке аз 6 зергурӯҳ функсияе мувофиқ гузошта мешавад:

$$F(s) = \sum_{k=1}^s p_k \quad (s = 1, \dots, m) \quad (2.2.2)$$

—аналоги дискретии функсияи тақсимот. Бигуздор T_1, T_2 — ҷуфти зергурӯҳи калимашаклҳои ихтиёрӣ аз маҷмуи $\{G\}$ ва

$$F^{(\alpha)}(s) = \sum_{k=1}^s p_k^{(\alpha)} -$$

функсияҳои дискретии мувофиқи онҳо, $\alpha = 1, 2$ ва $s = 1, \dots, 35$ бошанд.

Таърифи 1. Масофаи байни зергурӯҳи калимашаклҳои T_1 ва T_2 — ро адади мусбати $\rho(T_1, T_2)$ меномем, ки он бо формулаи зерин муайян карда мешавад:

$$\rho (T_1, T_2) = \sqrt{\frac{m}{2}} \max_s \left| \sum_{k=1}^s (p_k^{(1)} - p_k^{(2)}) \right|, \quad (2.2.3)$$

яъне масофаи байни бузургиҳои тасодуфии дискретӣ (бо маҷмуи якхелаи қиматҳои имконпазири $\bar{N}$) ҳамчун масофаи максималии байни функсияҳои дискретии онҳо $F^{(1)}(s)$ ва $F^{(2)}(s)$, ҳисоб карда мешавад, ки он бо коэффитсенти вазн $\sqrt{m/2}$ зарб карда шудааст» [2].

“Гипотезаи III. Гурӯҳи калимашаклҳое, ки садои хушоҳанг доранд «якҷинса» ва гурӯҳи калимашаклҳои ғайри онро (яъне душворталаффуз)-ро «ғайри якҷинса» меноманд.

Дар бораи «якҷинсагӣ»-и хушоҳангии гурӯҳи калимашаклҳо сухан ронда, мо шабоҳат, монандӣ, якнавъӣ ва наздикии онҳоро дар назар дорем.

Амсилаи риёзии III-гипотезаҳо. Бигузор γ – ягон адади мусбат бошад.

Таърифи 2. Гурӯҳи калимашаклҳои T_1 ва T_2 γ - якҷинса номида мешаванд (садои фораи доранд), агар масофаи байни онҳо

$$\rho (T_1, T_2) \leq \gamma, \quad (2.2.4)$$

ва γ - ғайри якҷинса номида мешавад (садои дағал доранд), агар масофаи байни онҳо

$$\rho (T_1, T_2) > \gamma \quad (2.2.5)$$

бошад.

Нобаробариҳои (2.2.4) ва (2.2.5) амсилаи риёзии гипотезаҳои III ном доранд.

Таърифи 3. γ -таснифгари аз як параметри ҳақиқии γ вобастаро алгоритми қабули қарор дар бораи тааллуқ доштани ҷуфти зергурӯҳи калимашаклҳои T_1 ва T_2 ба як ё ду сифатҳои гуногуни овозии калимашаклҳо меноманд.

Бигузор $\tau = \tau(\gamma)$ – суммаи миқдори ҳатогиҳои гипотезаи Н яқбора дар ду ҳолат бошад: иҷро нагардидани нобаробарии “якчинсагӣ” дар ҳолати ду зергурӯҳи калимашаклҳо ба як гурӯҳ мансуб будан ва иҷро нагардидани нобаробарии “ғайриякчинсагӣ” дар ҳолате, ки ду зергурӯҳи калимашаклҳо ба гурӯҳҳои гуногун мансуб бошанд. Онгоҳ барои γ – ҳои муайяншуда нишондиҳандаи иҷроиши гипотезаҳо бо бузургии π , аз рӯйи формулаи

$$\pi = 1 - \tau(\gamma)/L,$$

муайян карда мешавад, ки дар он L – шумораи масофаҳои мутақобилаи байни ҳамаи ҷуфтҳои зергурӯҳҳои калимашаклҳои маҷмуи **G** (дар ҳолати мо $L = G_6^2 = 15$) мебошад. Аз ин формула бармеояд, ки π метавонад қиматҳояшро аз порчаи $[0, 1]$ қабул кунад, илова бар ин $\pi = 0$ аст, агар $\tau = L (= 15)$ ва $\pi = 1$ аст, агар $\tau = 0$ бошад. Дар ҳолати аввал, гипотезаи Н бояд қорношоям эътироф карда шавад ва дар ҳолати дуюм - ба намунаи омӯзишӣ комилан мувофиқ мебошад” [4 - А].

“Аз сабаби он, ки самаранокии γ -таснифгар аз қимати параметри γ вобаста аст, зарур аст, ки чунин қимати онро пайдо намоем, ки барои он π қимати максималиро қабул мекунад. Моҳияти танзими γ -таснифгар аз рӯйи маълумоти интихобшудаи таълимӣ маҳз аз ҳамин иборат аст. Агар чунин танзимот қобили қабул бошад, пас дар бораи ҳалли масъалаи омӯзиши таснифгари γ ва моилияти он ба шинохти хушоҳангии калимашаклҳои тоҷикӣ аз маҷмуаҳои мухталиф суҳан рондан мумкин аст” [2], [3].

“Натиҷаҳои ниҳой дар мисоли коллексияи амсилаи G тавассути иҷрои пайдарпайи амалҳои зерин оварда мешаванд:

- ҳисобкунии симоҳои рақамӣ (басомади униграммаҳо) барои ҳамаи 6

зергурӯҳи калимашаклҳои маҷмуаи амсилаи G ;

- ҳисобкуниҳои масофаи понздаҳ ҷуфти $\rho(T_1, T_2)$ байни гурӯҳҳои маҷмуаи G бо формулаҳои (2.2.1), (2.2.2) ва (2.2.3) (натиҷаи ҳисобкуниҳо дар ҷадвали 2.2.2 оварда шудааст);

- ҳисобкунӣ бо ёрии алгоритми танзими γ -таснифгар интервали бехтарини қиматҳои γ , ки барои он бузургии $\tau = \tau(\gamma)$ суммаи миқдори ҳолатҳои ҳамогиҳои гипотезаи H қимати камтаринро соҳиб мегардад ва мувофиқан бузургии π нишондиҳандаи иҷрои гипотезаи H қимати максималиро қабул мекунад.

Ҷадвали 2.2.2 – Масофаи байни гурӯҳҳои маҷмуа ва зергурӯҳҳои санҷидашаванда

Матнҳо		Гурӯҳи калимашаклҳои душворталаффуз (G_1)			Гурӯҳи калимашаклҳои хушоҳанг (G_2)		
		G_{11}	G_{12}	G_{13}	G_{21}	G_{22}	G_{23}
Гурӯҳи калимашаклҳои душворталаффуз (G_1)	G_{11}						
	G_{12}	0.0423					
	G_{13}	0.1974	0.1863				
Гурӯҳи калимашаклҳои хушоҳанг (G_2)	G_{21}	0.4708	0.4580	0.6202			
	G_{22}	0.3399	0.3270	0.4893	0.1521		
	G_{23}	0.4455	0.4327	0.5949	0.1385	0.1074	

Дар ҷадвали 2.2.2 натиҷаҳои зерин ба даст омадаанд:

- маҷмуи ҳамаи ҷуфтҳои масофаҳо дар порчаи $[0.0423, 0.6202]$ ҷойгиранд, дар ин ҳолат масофаи хурдтарин дар байни зергурӯҳҳои G_{11} ва G_{12} дар гурӯҳҳои душворталаффуз ва масофаи калонтарин дар байни зергурӯҳҳои G_{13} дар гурӯҳи душворталаффуз ва G_{21} дар гурӯҳи хушоҳанг дида мешаванд;

- ниминтервали оптималии қиматҳои γ дар ҳудудҳои зерин аст

$$\gamma^{\text{опт}} \in [0.1974; 0.3270);$$

мувофиқи таърифи 3 ин маънои онро дорад, ки агар масофаи байни ду зергурӯҳи калимашаклҳо $\rho(T_1, T_2)$ аз қимати ниминтервали нишондодашудаи $\gamma^{\text{опт}}$ зиёд нашавад, онгоҳ ҷуфти зергурӯҳи калимашаклҳо ба ҳамон як гурӯҳ

тааллуқ доранд; агар зиёд бошад, онгоҳ ба гурӯҳҳои алоҳида мутааллиқанд;

- қайд мекунем, ки барои ҳамаи гурӯҳҳои маҷмуаи **G** гипотезаи III ва амсилаи риёзии он дар намуди таърифи 3 пурра тасдиқ гардид ва аз ин рӯ ҳосил шуд

$$\tau = \tau_{min} = 0,$$

яъне амсилаҳои математикии (2.2.4) ва (2.2.5) иҷро гардиданд;

- дар натиҷаи ин нишондиҳандаи самаранокӣ, кори амсилаи риёзии пешниҳодшудаи шинохти хушоҳангии калимашаклҳо баробар мебошад

$$\pi = \pi_{max} = 1.$$

Пас аз сохтани симои рақамии зергурӯҳҳои калимашаклҳо ва ҳисобкунии масофаҳои байни онҳо аз рӯи формулаи (2.2.3), чадвали масофаҳои ҳар як 6 гурӯҳи калимашаклҳои ин маҷмуа дар шакли зерин ҳосил гардид:

Чадвали 2.2.3 – Масофаи байни гурӯҳҳои маҷмуа ва зергурӯҳҳои санчидашаванда

Матнҳо		Гурӯҳи калимашаклҳои бетараф (G_3)		
		G_{31}	G_{32}	G_{33}
Гурӯҳи калимашаклҳои душворталаффуз (G_1)	G_{11}	0.1280	0.1155	0.1067
	G_{12}	0.1127	0.1187	0.0939
	G_{13}	0.2832	0.2610	0.2561
Гурӯҳи калимашаклҳои хушоҳанг (G_2)	G_{21}	0.3525	0.3592	0.3641
	G_{22}	0.2215	0.2283	0.2331
	G_{23}	0.3272	0.3339	0.3388

Дар чадвали 2.2.3 катакчаҳои сатри якум ва дуум ададҳои бо ранги ғафси моил нишон дода шудааст, ки ба масофаи ҳадди ақал аз объектҳои санчидашуда то элементҳои маҷмуаи матнҳо мувофиқат мекунанд. Натиҷаҳои бадастомада нишон медиҳанд, ки ҳамсояҳои наздиктарини сегонаи зергурӯҳҳои интиҳобшуда танҳо бо зергурӯҳҳои калимашаклҳои душворталаффуз якхела мебошанд.

Хулоса γ -таснифгар бо қимати доимии $\gamma = \gamma^{\text{опт}}$ дар гурӯҳҳои калимашаклҳои тасодуфан интихобгардида санчида шуд ва қобилияти шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ тавассути униграммаҳо 100%-ро ташкил дод” [4 – А, 5 – А, 6 – А].

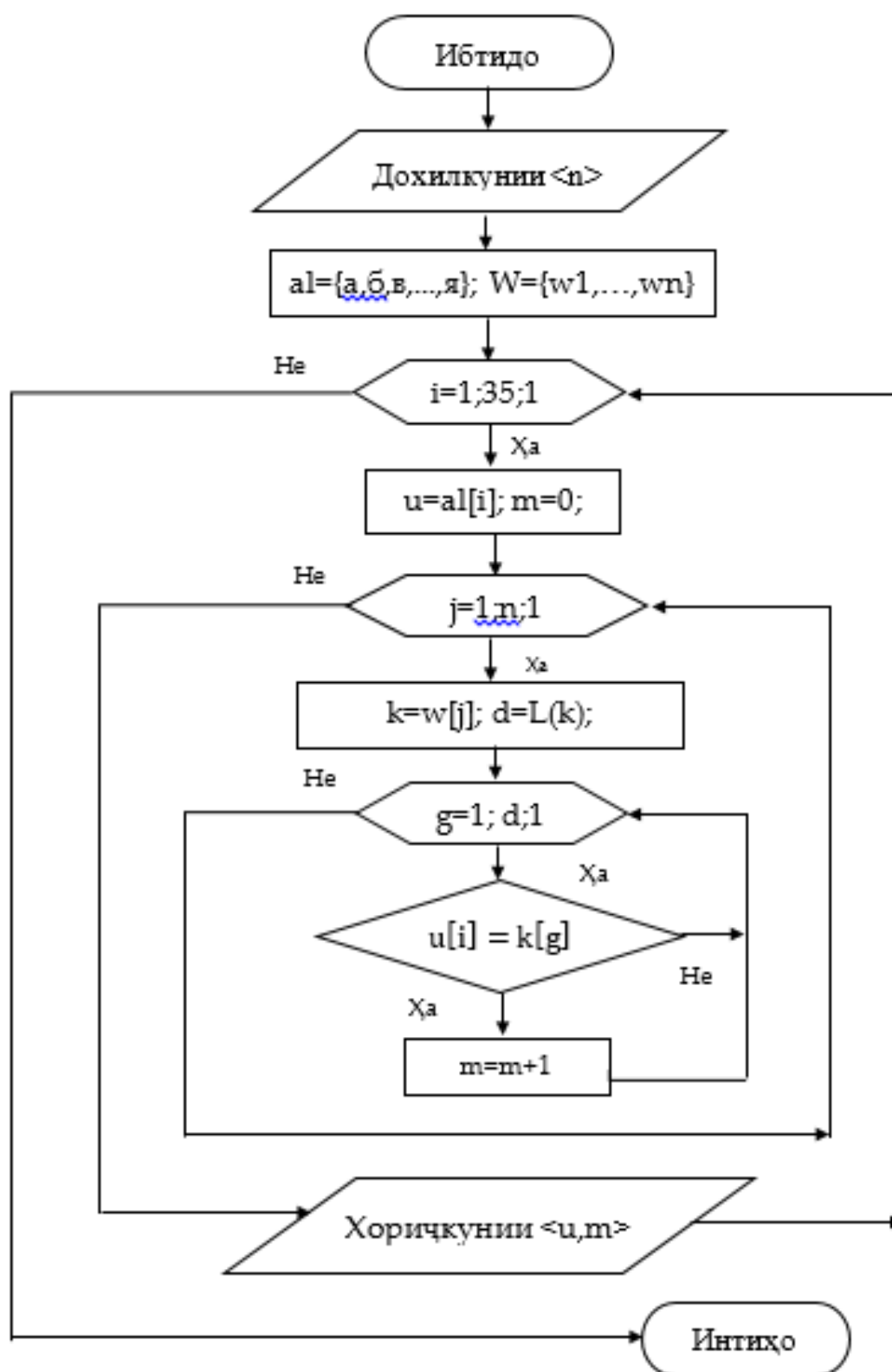
Дар **зербоби 2.3** амсилаи шинохти хушоҳангии калимашаклҳои тоҷикӣ ҳангоми истифодаи биграммаҳо (чуфти ҳарфҳои калимашакл) дида баромада шуд, ки дар ин ҳолат унсурҳои γ - таснифгарро биграммаҳои: **аа, аб, ав, ... , ая, ба, бб, ..., бя, ... , яя** ва симои рақамии онҳо иборат гардид. Се гурӯҳи калимашаклҳо аз маҷмуаи матнҳо интихоб гардида, ҳар кадоме аз онҳо дар навбати худ боз ба се зергурӯҳ чудо гардидаанд. Басомадҳои ҳар яке аз 1225 биграммаҳои зергурӯҳҳои калимашаклҳои аз маҷмуа интихоб гардида ҳисоб карда шуда, фосилаи байни онҳо ошкор ва байни ҳам муқоиса гардиданд. Самаранокии γ - таснифгар бо истифодаи биграммаҳо дида баромада шуда, амсилаҳои риёзӣ ва компютерии калимашаклҳо бо истифодаи биграммаҳо, ҳисоб карда шудаанд.

Дар **зербоби 2.4** амсилаи шинохти хушоҳангии калимашаклҳои тоҷикӣ ҳангоми истифодаи триграммаҳо (сегонаи ҳарфҳои калимашакл) дида баромада шуд, ки дар ин ҳолат унсурҳои γ - таснифгар аз триграммаҳои: **ааа, ааб, аав, ... , аая, аба, абб, ..., абя, ..., яая, яяб, ... , яяя** ва симои рақамии онҳо иборат гардид. Се гурӯҳи калимашаклҳо аз маҷмуаи матнҳо интихоб гардида, ҳар кадоме аз онҳо дар навбати худ боз ба се зергурӯҳ чудо гардидаанд. Басомадҳои ҳар яке аз 42875 триграммаҳои зергурӯҳҳои калимашаклҳои аз маҷмуа интихобгардида ҳисоб карда шуда, фосилаи байни онҳо пайдо ва байнан муқоиса гардидаанд. Самаранокии γ - таснифгар бо истифодаи триграммаҳо дида баромада шуда, амсилаҳои риёзӣ ва компютерии калимашаклҳо бо истифодаи триграммаҳо, ҳисоб карда шудаанд.

Дар зербоби 2.5 хулосаи боби дуюм оварда шудааст.

Дар боби 3 суҳан дар бораи сохтор, имконот ва татбиқи маҷмааи барномавии дар доираи диссертатсия омодагардида меравад. Маҷмааи дар доираи рисолаи илмӣ таҳиягардида “**ХУШСАДО**” номгузорӣ шудааст.

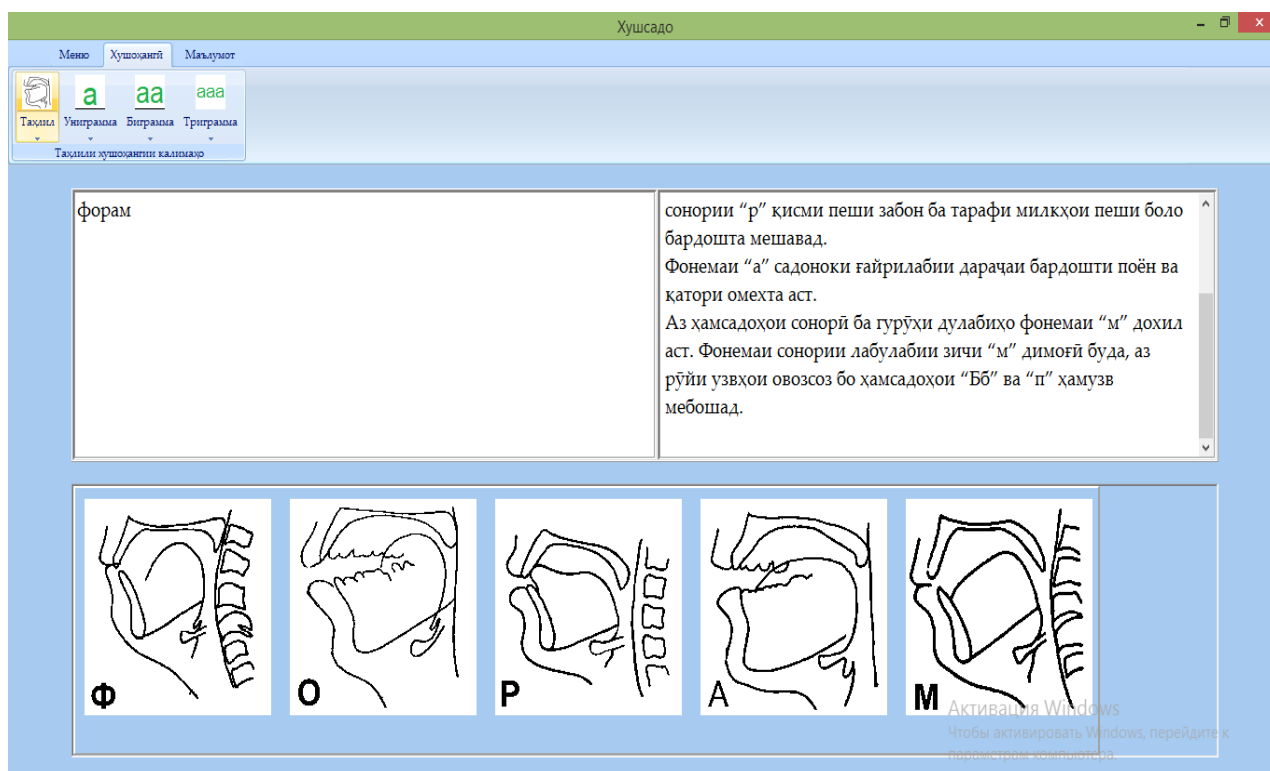
Дар зербоби 3.1 сохтори маҷмааи барнома ва қисматҳои он шарҳу тавзеҳ додашудааст. Алгоритми коркарди униграммҳои калимашаклҳои забони тоҷикиро дар шакли блок-схема тасвир менамоем (ниг. ба расми 3.1.1).



Расми 3.1.1 - Блок-схемаи коркарди униграммҳои калимашаклҳо

Дар зербоби 3.2 маҷмуаи матнии ҷамъоваришуда тавассути нармафзори таҳиягардида коркард шудааст.

Дар зербоби 3.3 дар асоси амсилаҳои математикии шинохти хушоҳангии калимашаклҳо, ки дар боби 2 оварда шуда буданд, амсилаи иттилоотӣ (компютерӣ)-и шинохти худкори сифати овозии калимашаклҳо таҳия шудааст. Барои таҳлили фонетикӣ ва рентгенограммавии калимаҳоро иҷро кардан фармони Таҳлилро аз нармафзори Хушсадо иҷро мекунем (ниг. ба расми 3.3.1).



Расми 3.3.1 – Таҳлили хусусиятҳои овозҳои калимашаклҳои зебо

Қисмати дигари барнома, ки ҳангоми гузариш ба речаи Васеъшуда пайдо мегардад барои ҳисобкунии симоҳои рақамии ҳарф, духарфӣ ва сеҳарфҳои гурӯҳи калимаҳо ва масофаҳои байни зергурӯҳҳои калимаҳо пешбинӣ шудааст (нигаред ба расми 3.3.2).

ХУЛОСА

Натиҷаҳои таҳқиқоти диссертатсионӣ, ҳалли нави масъалаи ҳамгироии мубрамро вобаста ба таҳияи асосҳои системаҳои шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ дар худ таҷассум мекунад. Дар асоси таҳқиқотҳои гузаронидашуда чунин натиҷаҳо ба даст оварда шудааст:

- ✓ таҳлили масъалаи хушоҳангии матн ва роҳҳои ҳалли он дар сарчашмаҳои мавҷудбуда ва таҳияи асосҳои андозагирии хушоҳангии калимашаклҳои забонҳои табиӣ, аз ҷумла забони тоҷикӣ [1-М, 2-М, 11-М];
- ✓ коркарди компютери матни Гулистони Саъдӣ, ошкор намудани маълумоти оморӣ ва хусусиятҳои ҳамнишинии овозҳои таркиби калимаҳои он [3-М];
- ✓ маҷмуаи матнии калонҳаҷм интихоб гардида, элементҳои он: униграмма, биграмма ва триграммаҳо аз нуқтаи назари математикӣ ва компютерӣ таҳқиқ гардидаанд [4-М, 5-М, 6-М];
- ✓ пайдо намудани аломатҳои душворталаффузӣ дар асосӣ пурсиши соҳибони забон, таснифи калимашаклҳои тоҷикӣ ва таҳия намудани барномаҳои компютери тавсифи аломатҳо [7-М, 8-М];
- ✓ ҳиҷои тоҷикӣ таҳқиқ гардида, алгоритм ва барномаи ҷудокунии ҳиҷоҳои калимаҳои тоҷикӣ таҳия карда шудааст [10-М, 12-М];
- ✓ барои ҳалли масъалаи шинохти худкори хушоҳангии калимашаклҳои забони тоҷикӣ маҷмааи барномавӣ дар забони барномасозии Delphi сохта шудааст [13-М].

ТАВСИЯҲОИ АМАЛӢ ДАР АСОСИ НАТИҶАҲОИ ТАҲҚИҚОТ

Натиҷаҳои таҳқиқотро барои ҳалли масъалаҳои забоншиносӣ ва адабиётшиносӣ, гузаронидани озмоишҳои гуногуни суҳансанҷии масъалаҳои мушаххас, системаҳои фонетикӣ ва системаҳои таълимию назоратӣ истифода бурдан мумкин аст. Натиҷаҳои ҳисобкунии компютерӣ нишон медиҳанд, ки барои таснифи калимашаклҳо ба синфҳо муайян кардани хусусиятҳои сифатӣ миқдории онҳо аҳамияти калони амалӣ дорад. Гуфта мешавад, ки

натиҷаҳои таҳқиқоти кори диссертатсионӣ барои хондани курсҳои махсус («Усулҳои таҳлили матн дар компютер», «Асосҳои лингвистикии тарҷумаи машинӣ», «Технологияҳои компютерӣ дар тадқиқотҳои забоншиносӣ»), иҷрои корҳои лабораторӣ ва мустақилона, курсӣ, дипломӣ ва магистрӣ барои донишҷӯён ва магистрҳои самтҳои «Информатика», «Забоншиносии компютерӣ» ва «Забоншиносӣ ва технологияҳои нави иттилоотӣ» дар донишгоҳҳои Ҷумҳурии Тоҷикистон, истифода мешавад.

Маҷмааи барномаи компютерии дар доираи диссертатсияи мазкур сохташуда барои истифодабарӣ дар соҳаҳои илму маориф, фарҳанг, забоншиносӣ, адабиётшиносӣ, системаҳои компютерии коркарди матн ва шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ, системаҳои зехни сунъӣ ва омӯзиши машинӣ барои ҳалли масъалаҳои мубрами забоншиносии компютерӣ пешбинӣ карда шудааст.

ФЕҲРИСТИ ИНТИШОРОТИ ИЛМИИ ДОВТАЛАБИ ДАРЁФТИ ДАРАҶАИ ИЛМӢ

а) Мақолаҳои, ки дар нашрияҳои тақризшавандаи Комиссияи олии аттестатсионии назди Президенти Ҷумҳурии Тоҷикистон ҷоп шудаанд:

[1 – М] Пиров С.М. К разработке основ эвфонометрии слов [Текст] / З.Д. Усманов, С.М. Пиров // Известия Академии наук Республики Таджикистан. Отделение физико-математических, химических, геологических и технических наук. - 2015. - №3 (160). - С. 39-42.

[2 – М] Пиров С.М. Дар бораи баъзе масъалаҳои хушоҳангии матнҳои тоҷикӣ [Матн] / С.М. Пиров // Паёми Донишгоҳи миллии Тоҷикистон. -2017, -№ 4-3. -С.105-109.

[3 – М] Пиров С.М. Таҳқиқи биграммаҳои “Тулистон”-и Саъдии Шерозӣ [Матн] / С.М. Пиров // Паёми Донишгоҳи миллии Тоҷикистон. -2017, -№ 1/5. - С.156-159.

[4 – М] Пиров С.М. Алгоритм распознавания благозвучия таджикских слов при использовании униграмм [Текст] / С.М. Пиров // Доклады Национальной академии наук Таджикистана. -2022, т. 65, № 1-2. С.57-62.

[5 – М] Пиров С.М. Алгоритми шинохти хушоҳангии калимаҳои тоҷикӣ бо истифодаи биграммаҳо [Матн] / С.М. Пиров // Паёми ДМТ, бахши илмҳои табиӣ. -2022. -№2. - С. 102-111.

[6 – М] Пиров С.М. Алгоритм распознавания благозвучия таджикских слов при использовании триграмм [Текст] / С.М. Пиров // Вестник, Филиала МГУ им. В.М.Ломоносов в г.Душанбе, раздел информатика. -2023, Том 1, -№1 (29). -С. 35 - 42.

б) Мақолаҳое, ки дар дигар навиришҳо ба таърифи расидаанд:

[7 – М] Пиров С.М. Дар бораи гурӯҳбандии калимаҳои забони тоҷикӣ аз рӯи талаффузашон [Матн] / С.М. Пиров // РТСУ. Перспективы развития фундаментальных и прикладных лингвистических исследований в Республике Таджикистан. – Душанбе: Персис, 2016. - 1(7). -С. 16 -18.

[8 – М] Пиров С.М. Дар бораи маҷмӯи аломатҳои тавсифкунандаи хушӯҳангии калимаҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи конференсияи илмӣ-амалии байналмилалӣ. ДТТ. -Душанбе -2017. -С. 352-355.

[9 – М] Пиров С.М. Басомади биграмаҳои ҳарфҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи Конференсияи ҷумҳуриявии илмӣ назариявии ҳаёати устодону кормандони ДМТ бахшида ба ҷашнҳои «5500-солагии Саразми бостонӣ», «700-солагии шоири барҷастаи тоҷик Камоли Хучандӣ» ва «20-солаи омӯзиш ва рушди фанҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илму маориф (солҳои 2020-2040)». – Ҷилди 1. -2020. -С. 39-40.

[10 – М] Пиров С.М. Алгоритми содаи ба ҳиҷоҳо ҷудокунии калимаҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи конференсияи байналмилалӣ илмӣ-амалӣ дар мавзӯи «Муаммоҳои муосири математика ва татбиқи он» бахшида ба «Солҳои 2020-2040 эълон гардидани 20-солаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ». -2022. -С. 158-161.

[11 – М] Пиров С.М. Таҳлили моделҳои баҳодиҳии хушӯҳангии матнҳо [Матн] / С.Ҳ. Мирзоев, С.М. Пиров // Маводи конференсияи байналмилалӣ илмӣ-амалӣ дар мавзӯи «Таҳлили компютери масъалаҳои илм ва технология» бахшида ба «Солҳои 2020-2040 эълон гардидани 20-солаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илму маориф» ва «75-солагии Донишгоҳи миллии Тоҷикистон». -2023. -С. 396-404.

**Шаҳодатномаҳо дар бораи бақайдгирии давлатии барномаи
компютерӣ барои МЭҲ:**

[12-М]. Пиров С.М. Шаҳодатнома. Ҷудокунии ҳиҷоҳои калимаҳои тоҷикӣ / С.М. Пиров // Свидетельство об официальной регистрации программы для ЭВМ в отделе по правам и защите прав автора Министерства культуры Республики Таджикистан (№ 150, 9.11.2022г.).

[13-М]. Пиров С.М. Автоматизированная система анализа текстов и оценка благозвучия таджикских слов [Текст] / С.Х. Мирзоев, А.А. Косимов, С.М. Пиров // Свидетельство о государственной регистрации информационного ресурса №1202500527 от 28.02.2025 внесен в реестр информационных ресурсов Республика Таджикистан.

Адабиёти истифодашуда

- [1].** Суханҳои ҳикматомӯзи Президент Эмомалӣ Раҳмон дар бораи забон. – Душанбе: Бунёди забон, 2015. – 116 сах.
- [2].** Усманов З.Д. Классификатор дискретных случайных величин –Доклады Академии наук Республики Таджикистан, 2017, т.60, № 7-8. – С. 291-300.
- [3].** Усманов З.Д. Алгоритм настройки кластеризатора дискретных случайных величин // Доклады АН РТ, 2017, т.60, № 9, – С. 392-397.

ТАДЖИКСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ

УДК: 004.5.7.9+519.25+811.21/.22

ББК: 22.18 (2Т)

П - 33

На правах рукописи



ПИРОВ САИДАЛИ МИРЗОАЛИЕВИЧ

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ КОЛИЧЕСТВЕННЫХ
КРИТЕРИЕВ ДЛЯ ОЦЕНКИ БЛАГОЗВУЧИЯ ТАДЖИКСКИХ СЛОВ**

А В Т О Р Е Ф Е Р А Т

**диссертации на соискание учёной степени кандидата технических
наук по специальности 05.13.11 – «Математическое и программное
обеспечение вычислительных машин, комплексов и компьютерных
сетей»**

Душанбе, 2025

Работа выполнена на кафедре «Информационные и коммуникационные технологии» механико-математического факультета Таджикского национального университета.

Научные руководители:

Усмонов З.Дж. - доктор физико-математических наук, академик НАНТ, профессор
Мирзоев С.Х. - доктор технических наук, и.о. профессора кафедры информатики Таджикского национального университета

Официальные оппоненты

Муминов Баходир Болтаевич - доктор технических наук, профессор, заведующий кафедрой искусственного интеллекта Ташкентского государственного экономического университета

Гафуров Миршафи Хамитович - кандидат технических наук, дотцент кафедры “Информационная технология и защита информации” Таджикского технического университета имени академика М.С. Осими

Ведущая организация

Институт математики имени А. Джураева НАНТ

Защита состоится «05» сентября 2025 года в «14:00» на заседании Диссертационного Совета 6D.KOA-049 при Таджикском техническом университете имени академика М.С. Осими по адресу: 734042, г. Душанбе, проспект академиков Раджабовых, 10.

С диссертацией можно ознакомиться в библиотеке Таджикского технического университета имени академика М.С. Осими и сайте [http: ttu.tj](http://ttu.tj).

Автореферат разослан «___» _____ 2025 года в соответствии с утверждённым реестром.

Ученый секретарь диссертационного совета 6D.KOA-049,
кандидат технических наук, доцент



Султонзода Ш.М.

Введение

Актуальность темы исследования. Актуальность темы исследования, прежде всего, связана с постановлением Правительства Республики Таджикистан «Об утверждении Программы использования и развития информационных технологий на таджикском языке», объявлением 2020-2040 годов «Двадцатилетием изучения и развития естественных, точных и математических наук в сфере науки и образования», а также с рядом принятых в этом направлении программ и стратегий.

История показывает, что во все времена ученые и исследователи уделяли особое внимание вопросу чистоты языка и сегодня продолжают эту главную обязанность. Защита языка от различных внешних дефектов, особенно сочетание и несогласие звуков словесного состава, является важнейшей задачей каждого индивидуума общества, в том числе творческих людей и интеллигенции, а обеспечение его красоты и чистоты — одним из требований благозвучия в науке анализа речи. Этот момент четко отмечен в книге Лидера нации «Мудрые слова Президента Эмомали Рахмона о языке. Уместно процитировать его здесь: «Содержать язык в чистоте, сохранять его подлинность и благозвучность – сыновний долг каждого из нас. Поэтому наши ученые и интеллигенты должны серьезно заняться исследованием важнейших вопросов языка, обновлять и создавать словари и энциклопедии, которые необходимы для повышения уровня знаний языка и культуры изучения языков нынешнего и будущих поколений» [1].

С появлением первых компьютерных программ появилась и идея обработки, хранения, анализа, произнесения и перевода текстов. Для обработки текстовой информации разработаны различные программы, решающие важнейшие задачи лингвистической науки. Одной из таких задач является задача благозвучия или удобочитаемости текстов и их оценка с помощью компьютера, которая решена для некоторых языках мира.

Например, в офисном пакете Office от корпорации Microsoft есть инструмент, который собирает статистику удобочитаемости и, помимо оценки лёгкости чтения и подсчёта количества сложных фраз, способен оценивать благозвучие.

Слово, выражение или текст называется благозвучным, если в его составе отсутствуют уникальные согласные, согласные из одного артикуляционного органа или из близких артикуляционных органов. В составе благозвучных слов гласных звуков больше, чем согласных и гласные звуки могут иметь до трёх открытых слогов. Для сохранения чистоты и благозвучия таджикского языка необходимо разработать стандарты, математические модели и специальные компьютерные системы с использованием всех средств, особенно современных информационных технологий. Данное исследование в основном сосредоточено на разработке основ системы автоматического распознавания эвфонии таджикских словоформ, основой которой является γ -классификатор.

Хотя критерии благозвучности не формализованы в некоторых языках, они существуют. Например, важнейшими критериями благозвучности текста являются: слоговая структура, соотношение низких и высоких звуков, доля звонких согласных, свистящих звуков и сочетание согласных. Вопрос красоты текста – очень сложный вопрос, и даже жители одного региона могут по-разному понимать красоту речи. Здесь необходимо учитывать множество факторов, на основе которых строится фундамент понимания каждого человека в отдельности. Кроме того, существует ряд особенностей, влияющих на качество информации из того или иного источника (например, в тексте может быть шрифт, качество бумаги в книге, голос и тон в речи).

Степень изученности и разработанности темы исследования.

Источники показывают, что первые мыслители Древней Греции начали изучать ораторское искусство и обсуждали его. Труд Аристотеля «Риторика» (Речь) является ярким доказательством этих утверждений. Относительно вопроса чистоты языка мыслители востока, такие как Абуали ибн Сина, Шамс Кайс Рази, Асади Туси, Шамс Гургони, Т. Зехни, Х. Шарифов, М. Шукуров, Б.

Камолиддинов и др. написали множество работ, которые считаются важнейшими источниками изучения различных сторон этого вопроса.

С начала прошлого века и по сегодняшний день разнообразные свойства фонетики современного таджикского литературного языка изучают известные учёные этой области И.И. Зарубин, Л. Бузургзода, В.С. Соколова, В.С. Расторгуев, Ш. Бобомуродов, Т.Н. Хаскашев, М. Файзов, А. Неъматов, Ш. Кабиров и их ученики. Профессор Т.Н. Хаскашев и М. Файзов внесли огромный вклад в изучение и развитие фонетики современного таджикского литературного языка. Этими отечественными исследователями проведены ценные научные, теоретические и практические анализы и сравнения в науке о таджикской фонетике, выявлены физиологические и акустические характеристики звуков таджикского языка с использованием специальных инструментов (рентгенограмма, политограмма и др.), используемых в области медицины и др.

Многие ученые, такие как Rudolf Flesch, Gunning R., Kincaid J., Spache G., Fry E., Hedman S., McLaughlin, G., В. Франчина, Р. Вакка, Оборнева И. В., Иванов К.В., Фомин А.Ю. и другие занимались изучением данного вопроса с применением математики и информационных технологий и разработали математические модели для оценки и анализа текстов на английском, итальянском, русском и других языках по различным параметрам и достигли значительных результатов, которые используются в крупнейших мировых компаниях, таких как Microsoft, национальных корпусах языков мира, географических и международных сайтах, различными министерствами и учреждениями.

В Республике Таджикистан в области обработки естественного языка, а именно таджикского языка, ученые Усмонов З.Дж., Исмоилов М.А., Тошхуджаев К.А., Умаров М.А., Зарипов С.А., Шарифов Ш., Гращенко Л.А., Худойбердиев Х.А., Солиев О.М., Давудов Г.М., Касимов А.А., Каюмов М.М. и другие проводят исследования вместе со своими учениками.

Создание компьютерного инструментария для автоматического распознавания качества звучания слов обеспечивает возможность составления эвфонометрического словаря, который может быть полезным дополнением к лингвистическому корпусу таджикского языка. Успехи в разработке основ эвфометрии слов неизбежно повлекут за собой ускоренное развитие систем автоматической оценки благозвучия предложений, текстов, произведений, учебников и творчества авторов.

Связь исследования с программами (проектами) или научными темами: Данное исследование выполнено в рамках реализации плана научно-исследовательской работы кафедры информационных и коммуникационных технологий механико-математического факультета Таджикского национального университета бюджетной темы «Применение математических методов и новых информационных технологий в сфере лингвистики и изучения методов алфавитного кодирования» (государственный регистрационный номер № 0122TJ14117 (01.01.2021-31.12.2025)) и исследовательского проекта «Применение методов моделирования и новых информационных технологий в сфере образования и лингвистики» (государственный регистрационный номер № 0116TJ00661 (01.01.2016-31.12.2021)).

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель работы. Разработка основ системы автоматического распознавания благозвучия словоформ таджикского языка и применение ее в виде комплекса программного обеспечения.

Задачи исследования. Чтобы достичь цели придётся решать следующие задачи:

- сбор и подготовка коллекции текстов для изучения и осуществления научных экспериментов разного типа, в том числе определения благозвучия словоформ;
- исследование информативности признаков, которые предназначены для

распознавания и определения благозвучия словоформ;

- проверка эффективности математических методов, которые, применяются при распознавании благозвучия словоформ;

- проектирование и применение комплекса программного обеспечения для автоматизированного различения эвфонии таджикских слов.

Объект исследования. В качестве объекта исследования определены математические и компьютерные модели анализа и оценки множества печатных словоформ таджикского языка.

Предмет исследования. Автоматическое распознавание благозвучия таджикских словоформ на основе признаков и частотности его N -грамм.

Методы исследования. Для решения целей и задач, поставленных в исследовании, используются фонетические методы, методы математической статистики, вычислительного эксперимента, теории множеств, системного анализа, распознавания, математическое и компьютерное моделирование и объектно-ориентированное программирование для разработки программных средств.

Теоретические основы исследования. Теоретическую основу диссертации составляют научные источники наших предков, научные результаты ученых разных эпох и времен, имеющие отношение к теме диссертации.

Научная новизна работы. Основные результаты, приведенные в диссертации, являются новыми, и они следующие:

- исследована информативность признаков, предназначенных для распознавания благозвучия словоформ;

- установлена эффективность использования N -граммы букв ($N=1, 2, 3$) в качестве признаков, определяющих благозвучия словоформ;

- показана эффективность применения классификатора Усманова З.Дж., который способен распознавать благозвучие коллекции словоформ с точностью до 96 %;

- впервые в Республике Таджикистане создан компьютерный

программный комплекс для распознавания благозвучия таджикских слов.

Положения, выносимые на защиту. Следующие положения будут выносимы на защиту:

1. Признаки, предназначенные для распознавания и определения благозвучия таджикских словоформ;
2. Количественный образ таджикских словоформ, характеризующих благозвучие в виде распределения частотностей буквенных N -грамм ($N=1,2,3$);
3. Эффективность применения классификатора З.Д. Усманова, как нового и соответствующего метода распознавания благозвучия таджикских словоформ;
4. Современный комплекс компьютерных программ для автоматического распознавания благозвучия таджикских словоформ.

Теоретическая значимость исследования. Теоретическая значимость работы состоит в том, что в ней экспериментально апробирован новый метод классификации дискретных случайных величин и установлена эффективность его применения для целей распознавания благозвучия печатных таджикских словоформ с буквенным алфавитом.

Практическая значимость исследования. Практическая значимость работы заключается в том, что она предназначена для реализации созданного в ней комплекса компьютерных программ в области науки и образования, культуры, языкознания, литературоведения и компьютерных систем обработки текстов для автоматического распознавания благозвучия таджикских словоформ.

Степень достоверности результатов диссертации. Полученные результаты подтверждаются серией вычислительных экспериментов по идентификации с достаточно высокой точностью различных коллекций текстов.

Соответствие диссертации паспорту научной специальности. Содержание исследования данной диссертации соответствует пунктам 1, 4 и 7

паспорта специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»:

П.1. Модели, методы и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования;

П.4. Системы управления базами данных и знаний;

П.5. Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения.

Личный вклад соискателя учёной степени. Постановка задачи и выбор методов исследования осуществлялись соискателем под руководством научных руководителей. Основные результаты диссертационной работы получены автором самостоятельно.

Апробация результатов работы. Материалы работы представлены и обсуждались на:

- научно-исследовательских семинарах Таджикского национального университета, Таджикского технического университета имени академика М.С. Осими, Технологического университета Таджикистана и Российско-Таджикского (Славянского) университета в период 2014-2023 гг.;

- международной научно-практической конференции «Перспективы развития науки и образования», 2016 г., Душанбе;

- международной научно-практической конференции «Роль ИКТ в инновационном развитии экономики Республики Таджикистан», 2017 г., Душанбе;

- международной научной конференции «Современные проблемы математики и их приложения», 14-15 июня 2017 г., Душанбе, Куляб;

- международной научно-практической конференции «XII Ломоносовские чтения», посвященной Дню таджикской науки и 30-летию установления дипломатических отношений между Республикой Таджикистан

и Российской Федерацией (29-30 апреля 2022). Часть I. Естественные науки. – Душанбе, 2022;

- международной научно-практической конференции “Современные проблемы математики и её приложения”, посвященной 20-летию развития естественных, точных и математических наук 2020-2040 годы. -Душанбе - 2022;

- международной научно-практической конференции «Компьютерный анализ проблемы наука и технология» посвященной «2020-2040 годы объявлены двадцатилетием изучения и развития естественных, точных и математических наук» и «75-летию Таджикского национального университета». -Душанбе -2023.

Публикации по теме диссертации. По теме диссертации опубликовано 13 работ, из них 6 статей в журналах из перечня, рекомендованного ВАК при Президенте Республики Таджикистан, [1-А - 13-А].

Структура и объём диссертации. Диссертация состоит из введения, трех глав, заключения и библиографического списка. Библиографический список включает 137 наименований. Основная часть диссертации изложена на 143 страницах. Диссертация содержит 59 рисунков и 13 таблиц.

Автор выражает свою особую признательность и благодарность научными руководителям – доктору физ.-мат. наук, профессору, академику НАНТ, глубокоуважаемому Усманову З.Д. и доктору технических наук, доценту Мирзоеву С.Х., а также сотрудникам Таджикского национального университета, Института математики имени А. Джураева НАНТ и Таджикского технического университета имени академика М.С. Осими.

ОСНОВНАЯ ЧАСТЬ ИССЛЕДОВАНИЯ

Введении диссертации включает в себя обоснование актуальности выбранной темы диссертации, определение цели, задач, объекта и предмета исследования, раскрытие новизны, положения, вынесенные на защиту, изложение теоретической и практической значимости работы, описание

методологической базы и материал исследования, соответствие диссертации паспорту научной специальности, сведения об апробации работы и ее структуре.

В главе 1 будет рассматриваться проблема благозвучия текстов, анализируются существующие методы её решения, аспекты появления звуков речи и их физические и фонологические характеристики.

В подглаве 1.1 исследованы мнения исследователей разных эпох и времен из области филологии, математики и информационных технологий о вопросе благозвучия и трудного произношения слов.

В подглаве 1.2 рассматриваются аспекты появления звуков речи. Фонетический строй языков человеческого общения имеет очень сложную природу, поэтому необходимо изучать различные его аспекты. Наша речь, прежде всего, является акустическое явлением и имеет физическую природу. Каждое акустическое событие происходит в определенное время и с силой и скоростью, свойственным ему.

На физической и акустической основе анализируются и исследуются физическая природа звуков речи и их характеристики: длина, уровень тона и силы, формантная и спектральная структура. Исследуются и сравниваются эти физические признаки звуков речи в разных фонетических состояниях и ситуациях, а также определяется степень владения ими этими характеристиками по сравнению с другими звуками. На этих признаках основана и акустическая классификация звуков.

Звуки, которые мы слышим, имеют источник и причину появления. Источником и причиной появления звуков речи является речевой аппарат человека и работа его органов (см. рисунок 1.2.1).

Основной частью гортани, играющей важную роль в звукообразовании, являются голосовые связки. Голосовые связки способны вибрировать примерно от 40 до 1700 раз в секунду, но при нормальной речи их пульсация равна 85-200 (в мужской речи) или 160-340 (в женской речи) колебаний в

секунду. Мужские голосовые связки толще, а женские тоньше, поэтому количество голосовых колебаний у мужчин ниже, а у женщин выше.

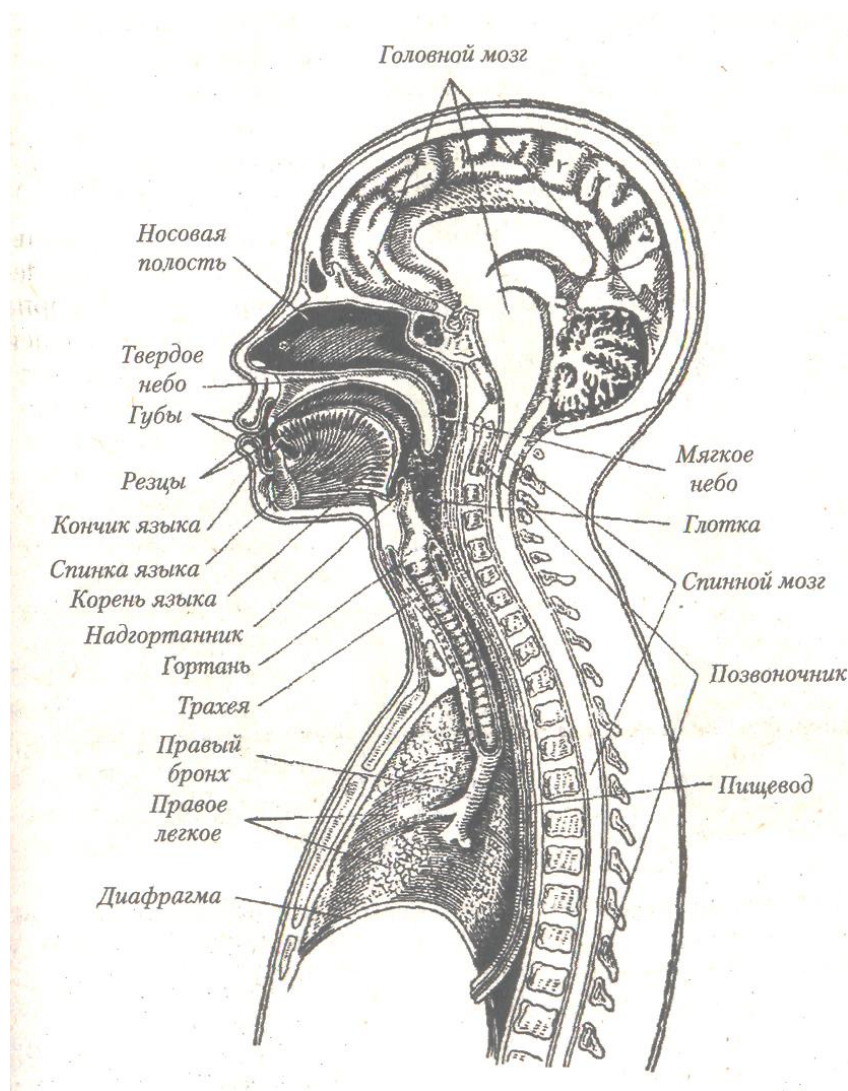


Рисунок 1.2.1 - Речевой аппарат

Каждый язык имеет свою систему звуков, которые создаются с помощью разных органов. Активными голосовыми органами таджикского языка являются: голосовые связки, стенки гортани, мягкое небо с языком, язык и губы, которые в большей степени способствуют образованию звуков. Например, из 30 фонем (звуков) таджикского языка 20 фонем (66,7%) находятся в передней части языка и только 10 других фонем: гласные у, ў, о и согласные г, к, х, ғ, қ, ъ, ҳ (33,3%) создаются в задней части языка. К пассивным голосовым органам относятся твердое небо, зубы, десны и полость носа.

Хотя особенности речевого аппарата таджикского языка еще не исследованы, о некоторых из них имеются предварительные сведения. Одной из таких особенностей, принадлежащих голосовой системе таджикского языка, является музыкальная особенность – благозвучие. Благозвучие – это идеальное сочетание голосов, которое приносит легкость говорящему и удовольствие слушателю. Требования к произношению должны быть конкретными в соответствии с фонетическими характеристиками языка. Благозвучие всегда зависит от фонетического разнообразия национального языка.

В **подглаве 1.3** рассматриваются физические свойства звуков, особенно звуки таджикского языка.

В **подглаве 1.4** речь идёт об артикуляционной характеристике таджикских звуков. Современный алфавит таджикского литературного языка состоит из 35 букв и 30 звуков (фонем), которые делятся на две основные группы: 1) гласные и 2) согласные. В произношении гласных звуков никаких препятствий нет. Кроме того, гласные отличаются от согласных и другими характеристиками. Поскольку в произношении гласных нет никаких препятствий, можно сказать, что это приятные звуки. Согласных относительно больше, и они играют важную роль в организации слов и текста. При произношении согласных они сталкиваются с различными препятствиями, которые в той или иной степени становятся на пути голосового потока и вызывают затруднения в произношении. Проблема нарушения благозвучия словоформ зависит и от количества согласных. Чтобы внести ясность в то, о чем мы говорим, приведем аппарат органов речи, полученный с помощью аппарата рентгенограммы (см. рисунок 1.4.1).

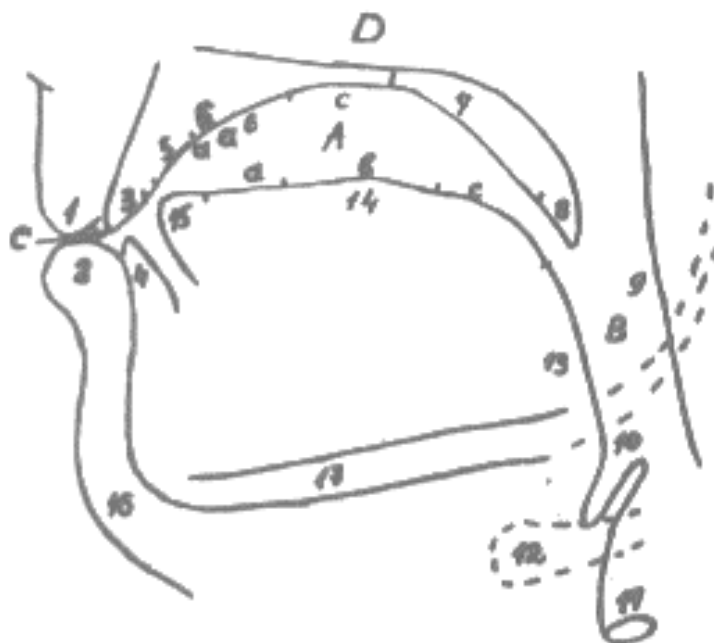


Рисунок 1.4.1 - Отображение органа речи на рентгенограмме

На рентгенограмме, приведенной на рисунке 1.4.1 органы произношения обозначены цифрами, а полости речевого аппарата – буквами. А–Полость рта, Б – полость глотки, С –полость дверца рта, Г –полость носа. Цифрами указаны: 1 – верхняя губа; 2 – нижняя губа; 3 – верхние зубы; 4 – нижние зубы; 5 – голосовые связки; 6 – твердое нёбо: передняя часть, средняя часть и задняя часть; 7 – мягкая нёбо; 8 – язычок; 9 –стенка глотки; 10 – надгортанник; 11 — гортань;12 – подъязычная кость; 13 – корень языка; 14 — спинка языка: передняя часть, средняя часть и задняя часть; 15 – кончик языка; 16 – подбородок; 17 – нижняя челюсть.

Подглава 1.5 первой главы включает в себя выводы, сделанные по итогам данной главы.

Глава 2 посвящена исследованию вопросов формирования основ эвфонометрии таджикского языка, использования математических и компьютерных моделей с применением элементов униграмм, биграмм и триграмм.

В **подглаве 2.1** рассматривается формирование основ эвфонометрии таджикского языка. «Эвфония (благозвучие) – понятие *качественное*, ибо

всякий индивид по-своему оценивает звучание слова, предложения, стихотворения, текста и речи. Вместе с тем рассматриваемое понятие – *статистическое*, поскольку совокупность индивидов относительно одних и тех же объектов во многих случаях высказывает “общее” мнение. Этот факт служит основой для развития эвфонометрии – новой научной дисциплины, нацеленной на изучение закономерностей благозвучия письменной и устной речи математико-статистическими методами. Практическая значимость таких исследований заключается в несомненном влиянии качества звучания речи на результаты восприятия информации. Первый этап разработки основ эвфонометрии начинается с создания или же с обращения к уже имеющейся **коллекции текстов** (ещё лучше к лингвистическому корпусу) того или иного естественного языка. Коллекция должна быть достаточно представительной, включать в надлежащих пропорциях художественную литературу, газетные и журнальные статьи, научные и научно-популярные тексты из разных областей знаний, общественно-политические произведения, драматургию и т.п. Желательно, чтобы объём коллекции исчислялся десятками миллионов словоупотреблений» [1 – А].

На основе 20 произведений персидско-таджикских писателей разных эпох и произведений разных жанров мы составили коллекцию текстов, состоящую из 6811 электронных страниц и содержащую 2206587 слов. Статистические результаты данного исследования представлены в таблице 2.1.1.

Таблица 2.1.1 - Статистика исследуемой коллекции текстов

№	Номи адиб	Номи асар	Микдори саҳифа	Микдори калима
1	Абдуррамони Ҷомӣ	Баҳористон	139	31 934
2	Саъдии Шерозӣ	Гулистон	150	39 589
3	Носири Хисрав	Сафарнома	100	34 523
4	Ибни Муқаффаъ	“Калила ва Димна”	120	39 647
5	Садриддин Айний	«Марги судхӯр»	220	57 926
6	Ҳусайн Воизи Кошифӣ	Фугувватнома	336	93 874
7	Мухаммадиев Ф.	Шаби савум	288	94 952

Продолжения таблица 2.1.1

8	Меҳмон Бахтӣ	Тилисми Сайхун	45	12 020
9	Сорбон	Романи Туғрал	190	57 876
10	Муҳаммадҷони Шакурии Бухорой	Садри Бухоро	188	112 709
11	Муҳаммадҷони Шакурии Бухорой	Хуросон аст ин чо	361	92 104
12	Чалолуддини Балхӣ	800 ғазали дилошӯб	287	83 215
13	Эмомалӣ Раҳмонов	Нигоҳе ба таърих ва тамаддуни ориёӣ	371	144 387
14	Ғафуров Б.Ғ.	Тоҷикон	870	305 673
15	Лоик Шералӣ	Куллиёт	830	210 713
16	Абулқосим Фирдавсӣ	Шоҳнома. Ҷилди I	277	57 694
17	Худойназар Асозода	Достони зиндагӣ (китоби 4-ум)	321	145 205
18	Худойназар Асозода	Достони зиндагӣ (китоби 5-ум)	480	223 087
19	Абулфазли Байҳақӣ	Таърихи Байҳақӣ	756	235 890
20	Маҷлиси Олии ҶТ	Кодекси андоз	482	133 569
			6811	2 206 587

На основе результатов опроса общественного мнения, проведенного посредством анкеты, словоформы были разделены на три класса. Респонденты ставили справа от каждой словоформы из списка $\bar{L}$ одну из трех цифр: -1, +1 и 0, которые кодируют качественные признаки, соответствующие звукам «трудный», «приятный» и «средний». Выбранные словоформы мы предъявляли экспертам (лингвистам), они присваивали каждому слову одну из трех цифр: -1, +1 или 0, соответствующих знаков качества произношений «трудный», «приятный» и «средний» для их кодирования. Экспертам была представлена анкета объемом шесть печатных страниц в виде таблицы.

Эксперт исследует словоформы в анкете, если она трудна в произношении, то справа от нее ставит цифра -1, если произношение словоформы не сложное, не приятное, а среднее, то цифра 0 – справа от него, и если произношение слова приятное, то рядом пишет цифру +1.

На основе различных особенностей звуков таджикского языка, закономерностей их сочетания, опроса носителей языка и компьютерной обработки словоформ мы выявили 16 признаков, описывающих различное звуковое качество словоформ таджикского литературного языка. Признаки,

выбранные для количественного и качественного описания благозвучия словоформ, представлены в таблице 2.1.2 с их возможными значениями.

Таблица 2.1.2 - Признаки количественных и качественных характеристик словоформ

№	Признак словоформ	Возможные значения	Число позиции признака
1	Длина (количество букв)	От 1 до 30	2
2	Число гласных	От 1 до 15	2
3	Число согласных	аз 1 то 15	2
4	Согласный Ғ	От 1 до 15	1
5	Согласный Ж	1-да, 0-нет	1
6	Согласный Қ	1-да, 0-нет	1
7	Согласный Ҳ	1-да, 0-нет	1
8	Согласный Ч	1-да, 0-нет	1
9	Согласный Ъ	1-да, 0-нет	1
10	Биграмма НГ	1-да, 0-нет	1
11	Биграмма НД	1-да, 0-нет	1
12	Биграмма РД	1-да, 0-нет	1
13	Биграмма СТ	1-да, 0-нет	1
14	Биграмма ФТ	1-да, 0-нет	1
15	Биграмма ХТ	1-да, 0-нет	1
16	Биграмма ШТ	1-да, 0-нет	1

В подглаве 2.2 рассмотрена модель распознавания благозвучия таджикских словоформ с использованием униграмм. В этой и последующих подглавах данной главы используется системный подход, который был предложен в 2015 году для разработки основ эвфонометрии естественных языков, в том числе таджикского языка, которой может быть применен к различным естественным языкам. В основе этого подхода лежит γ -классификатор дискретных случайных величин, разработанный академиком НАНТ Усмоновым З.Дж. Упомянутый классификатор может быть использован для решения различных задач лингвистической науки и выявления характеристик различных текстовых единиц, в том числе для определения произношения словоформ. В этой части работы для определения звукового качества словоформ мы используем упомянутый выше классификатор, который представляет собой математическую триаду, состоящую из трех частей:

- цифровой портрет (ЦП), т.е. количественное описание словоформ;
- формулы расстояния между ЦП словоформы;
- алгоритм принятия решения на основе образов.

Мы используем этот классификатор для автоматического распознавания голосовых качеств определенного набора словоформ: поэтического или прозаического произведения, заранее выбранного из корпуса текстов. Корпус текстов – это совокупность текстов разных жанров, хранящихся в электронном виде. Корпус текстов используется для проведения различных исследований в области языкознания, истории и других областях. Мы извлекли список некоторых слов из коллекции таджикских текстов. В таком списке собраны наиболее часто употребляемые словоформы «грубых», «нейтральных» и «благозвучных» звуков. Из коллекции таджикских текстов извлекаем список некоторых слов. Такой список содержит наиболее употребительные словоформы “грубого”, “нейтрального” и “приятного” звучаний.

Множество таджикских словоформ обозначаем буквой G и разделим его по качеству звучания на 3 группы: G_1 – грубозвучные, G_2 – благозвучные и G_3 – нейтральные слова. Из каждой группы выбираем по 1000-1500 словоформ. Группу G_1 разделим каким-либо образом на 3 подгруппы словоформ – G_{11} , G_{12} , G_{13} – приблизительно одинаковых размеров словоформ (в смысле числа словоформ). Далее, используя какой-либо способ, также разделим группу G_2 на 3 подгруппы словоформ – G_{21} , G_{22} , G_{23} , которые по количеству словоформ являются приблизительно одинаковыми. Подгруппы G_{1i} и G_{2j} будем рассматривать как подгруппы словоформы (тексты).

«В качестве элементов для описания количественного образа словоформы используем единый таджикский алфавит, состоящий из 35 букв (униграмм):

а, б, в, ..., э, ю, я.

Для каждой из 6-и подгруппы G_{1i} и G_{2j} вводим данные единого цифрового портрета, который задан в виде формулы закона распределения.

$$\begin{array}{lcl} \bar{N} & : & 1, \quad 2, \quad . \quad . \quad . \quad , 35 \\ P & : & p_1, \quad p_2, \quad . \quad . \quad . \quad , p_m, \end{array} \quad (2.2.1)$$

в которой первая строка обозначает занумерованный в алфавитном порядке список $m = 35$ таджикских букв (униграмм), а во второй строке их частотность для конкретной подгруппы, причём

$$\sum_{k=1}^{35} p_k = 1.$$

Каждой из 6 подгруппе словоформы соответствует функция:

$$F(s) = \sum_{k=1}^s p_k \quad (s = 1, \dots, m) \quad (2.2.2)$$

– дискретный аналог функции распределения. Предположим, что произвольная пара подгруппы словоформ T_1 и T_2 из $\{G\}$ и

$$F^{(\alpha)}(s) = \sum_{k=1}^s p_k^{(\alpha)}$$

являются соответствующими им дискретными функциями, $\alpha = 1, 2$ и $s = 1, \dots, m$.

Определение 1. Положительное число $\rho(T_1, T_2)$, которое определяется формулой

$$\rho(T_1, T_2) = \sqrt{\frac{m}{2}} \max_s \left| \sum_{k=1}^s (p_k^{(1)} - p_k^{(2)}) \right|, \quad (2.2.3)$$

будем называть расстоянием между подгруппы T_1 и T_2 , иными словами, расстояние между дискретными случайными величинами (с одинаковым

набором $\bar{N}$ возможных значений) находятся как максимальное расстояние между их дискретными функциями $F^{(1)}(s)$ и $F^{(2)}(s)$, умноженное на весовой коэффициент $\sqrt{m/2}$ » [2].

«ГИПОТЕЗА III. Слова, которые благозвучны, – “однородны”, а которые неблагозвучны сладко (то есть, звучат грубо) – “не однородны”.

Когда говорим об “однородности” благозвучия слов, имеется в виду, что они похожие, одинаковые, однотипные, родственные и т.п.

Математическая модель III-гипотезы. Пусть γ есть некоторое положительное число.

Определение 2. Подгруппы T_1 и T_2 называются γ - однородными (звучат сладко), если

$$\rho(T_1, T_2) \leq \gamma, \quad (2.2.4)$$

и γ - неоднородными (звучат грубо), если

$$\rho(T_1, T_2) > \gamma, \quad (2.2.5).$$

Соотношение (2.2.4) и (2.2.5) можно рассматривать в качестве математической модели гипотезы III.

Определение 3. γ - классификатор – зависящий от одного вещественного параметра γ алгоритма принятия решения об отнесении пары подгруппы T_1 и T_2 к одному или двум разным качествам звучания слов.

Пусть $\tau = \tau(\gamma)$ – суммарное количество нарушений гипотезы III одновременно в двух случаях: невыполнения неравенства “однородности” в случае двух подгруппы словоформ, принадлежащих одной группе, и невыполнения неравенства “неоднородности” в случае двух подгруппы словоформ, принадлежащих разным группам. Тогда для фиксированного γ

показатель выполнения гипотезы будет определяться значением π , задаваемым следующей формулой

$$\pi = 1 - \tau(\gamma)/L,$$

в которой L – число взаимных расстояний между всеми парами подгруппы словоформ из модельной коллекции $\mathbf{G}$ ($L = G_6^2 = 15$). Из приведенной формулы следует, что величина π может принимать свои значения из отрезка $[0,1]$, причём $\pi = 0$, если $\tau = L$, и $\pi = 1$, если $\tau = 0$. В первом случае гипотезу III следует признать непригодной, а во втором случае гипотезу III следует признать полностью согласованной с обучающей выборкой» [4 - А].

«Так как эффективность γ - классификатора имеет зависимость от значения параметра γ , следует найти такое его значение, при котором величина π принимает своё максимальное значение. Именно в этом и состоит суть настройки γ - классификатора на данных обучающей выборки. Если такая настройка будет приемлемой, то можно говорить о решении задачи обучения γ - классификатора и его предрасположенности к распознаванию благозвучия таджикских словоформ разных коллекций текстов» [2], [3].

«Заключительные результаты на примере коллекции G приведены путём последовательного выполнения следующих действий:

- вычисление цифровых портретов для всех шести подгрупп словоформ модельной коллекции $\mathbf{G}$;
- вычисление пятнадцати парных расстояний $\rho(T_1, T_2)$ между подгруппами модельной коллекции $\mathbf{G}$ по формулам (2.2.1), (2.2.2) и (2.2.3) (результаты расчетов показаны в таблице 2.2.2);
- вычисление посредством «алгоритма настройки γ - классификатора оптимального интервала значений γ , для которого величина $\tau = \tau(\gamma)$ суммарного числа случаев нарушения гипотезы III достигает минимального

значения и, следовательно, величина π показателя выполнения гипотезы III принимает максимальное значение.

Таблица 2.2.2 – Расстояния между подгруппами коллекции G

Подгруппы		Группа грубозвучных словоформ (G_1)			Группа сладкозвучных словоформ (G_2)		
		G_{11}	G_{12}	G_{13}	G_{21}	G_{22}	G_{23}
Группа грубозвучных словоформ (G_1)	G_{11}						
	G_{12}	0.0423					
	G_{13}	0.1974	0.1863				
Группа сладкозвучных словоформ (G_2)	G_{21}	0.4708	0.4580	0.6202			
	G_{22}	0.3399	0.3270	0.4893	0.1521		
	G_{23}	0.4455	0.4327	0.5949	0.1385	0.1074	

В таблице 2.2.2 получены следующие результаты:

- совокупность всех пар расстояний размещается на отрезке $[0.0423, 0.6202]$, при этом минимальное расстояние реализуется между подгруппами G_{11} и G_{12} в грубозвучных подгруппах, а максимальное – между G_{13} в грубозвучных и G_{21} в сладкозвучных подгруппах;
- оптимальный полуинтервал значений γ оказывается в пределах

$$\gamma^{\text{опт}} \in [0.1974; 0.3270);$$

согласно определению 3 это значит, что если расстояние $\rho(T_1, T_2)$ между двумя подгруппами словоформ не превышает значения $\gamma^{\text{опт}}$ из указанного полуинтервала, то пара подгруппы относятся к одной и той же группе; если же превышает, то относятся к разным группам словоформ;

- отметим, что для всех подгруппы модельной коллекции G полностью подтвердилась гипотеза III и её математическая интерпретация в виде определения 3, и потому получено

$$\tau = \tau_{\min} = 0,$$

иными словами, неравенств (2.2.4) и (2.2.5) выполнены;

- вследствие чего показатель эффективности, предложенной в

диссертационной работе математической модели распознавания благозвучия словоформ оказался равным

$$\pi = \pi_{max} = 1.$$

После того, как был сформирован ЦП подгруппы словоформ и вычислены расстояния по формуле (2.2.3), получена следующая таблица расстояний от каждого из подгруппы до всех 6 подгрупп исходной модельной коллекции:

Таблица 2.2.3 – Расстояния между подгруппами коллекции и тестируемыми подгруппами

Группы		Группа нейтральных слов (G_3)		
		G_{31}	G_{32}	G_{33}
Группа грубозвучных словоформ (G_1)	G_{11}	0.1280	0.1155	0.1067
	G_{12}	0.1127	0.1187	0.0939
	G_{13}	0.2832	0.2610	0.2561
Группа сладкозвучных словоформ (G_2)	G_{21}	0.3525	0.3592	0.3641
	G_{22}	0.2215	0.2283	0.2331
	G_{23}	0.3272	0.3339	0.3388

В таблице 2.2.3 выделены пары ячеек, которые соответствуют минимальным расстояниям от тестируемых объектов до элементов модельной коллекции текстов. Результаты показывают, что ближайшими соседями выбранной тройки подгрупп оказались однородные только с подгруппами грубозвучных слов.

Заключение. Проведено тестирование γ -классификатора, имеющего фиксированное значение $\gamma = \gamma^{\text{опт}}$, на случайных выборках группы, по результатам которого подтверждена 100 % -ная способность распознавания благозвучия таджикских слов» [4 – А, 5 – А, 6 – А].

В подглаве 2.3 рассмотрена модель распознавания благозвучия таджикских слов при использовании биграмм, где в качестве элементы γ -классификатора выступали биграммы: **аа, аб, ав, ... , ая, ба, бб, ..., бя, ... , яя** и их ЦП. Из коллекции G выбрано 3 группы словоформ: G_1 – грубозвучные, G_2 – сладкозвучные и G_3 – нейтральные слова и каждое из них в своей очереди

разделены на 3 подгруппы. Из каждой группы выбираем по 1000-1500 слов. Для каждой подгруппы сопоставлен цифровой портрет – распределения частотностей символьных биграмм.

Исследована эффективность γ - классификатора с использованием биграмм, рассчитаны математические и компьютерные модели словоформ с использованием биграмм.

В **подглаве 2.4** рассмотрен метод распознавания созвучности таджикских словоформ при использовании триграмм (триад словообразных букв), в данном случае элементы γ -классификатора взяты из триграмма: **ааа, ааб, аав, ..., аяа, аба, абб, ..., абя, ..., аяа, ааб, ..., яяя** и их цифровой портрет. Из коллекции текстов были выбраны три группы словоформ, каждая из которых далее разделена на три подгруппы. Были подсчитаны частоты каждой из 42875 триграмм выбранных из коллекции подгрупп словоформ, найдено расстояние между ними и проведено их сравнение. Была исследована эффективность γ -классификатора с использованием триграмм, а также рассчитаны математическая и компьютерная модели словоформ с использованием триграмм. В **подглаве 2.5** подводится итог второй главе.

В **главе 3** рассказывается о структуре, возможностях и реализации программного комплекса, подготовленного в рамках диссертации. Комплекс, разработанная в рамках научной диссертации, носит название «ХУШСАДО». Структура программы и ее части описаны в **подглаве 3.1**.

Опишем алгоритм обработки униграмм таджикских словоформ в виде блок-схемы (см. рисунок 3.1.1).

В **подглаве 3.2** собранная коллекция текстов обрабатывается разработанным программным обеспечением.

В **подглаве 3.3** на основе математических моделей распознавания благозвучие словоформ, представленных в главе 2, разработана информационная (компьютерная) модель автоматического распознавания голосовых качеств словоформ. Для фонетического анализа слов и

отображение рентгенограммы фонемы нужно выполнить команду «Тахлил» из программы «ХУШСАДО» (см. рисунок 3.3.1).

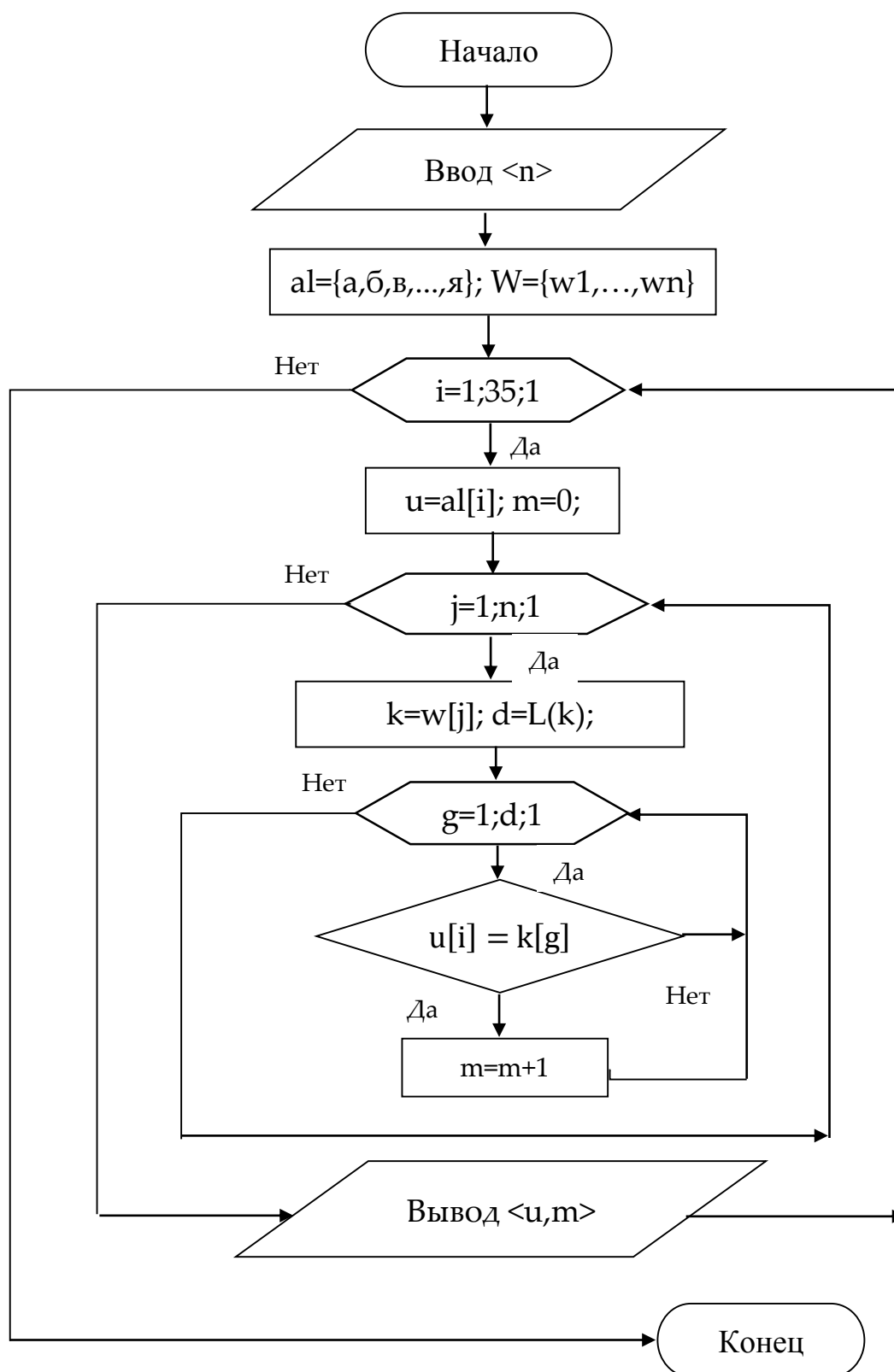


Рисунок 3.1.1 - Блок-схема обработки униграммы словоформы

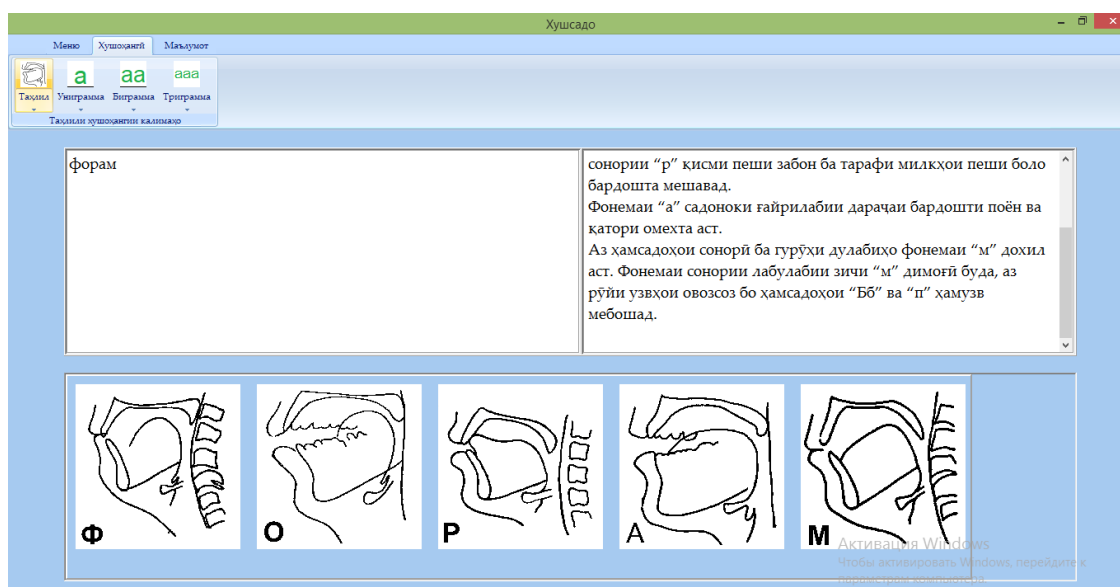


Рисунок 3.3.1 – Анализ голосовых характеристик красивых словоформ

Другая часть программа **Хушсадо**, которая появляется при переходе на режиме **Расширенный (Васеъшуда)** и служить для вычисления цифровых портретов (ЦП) униграмм, биграмм, триграмм группы словоформ и их расстояний (см. рисунок 3.3.2).

№	Уни	Бас.м	Бас.нис.	Чамғи бас.	№	Уни	Бас.м	Бас.нис.	Чамғи бас.	№	Уни	Бас.м	Бас.нис.	Чамғи бас.	$D(n,m)= F(n)-F(m) $	Dm
1	а	340	0,1318340	0,13183404	1	а	496	0,1656090	0,16560934	1	а	519	0,1624920	0,162492	0,00311717609081302	0,00311717609081302
2	б	101	0,0391624	0,17099651	2	б	110	0,0367270	0,20233722	2	б	83	0,0259866	0,188478	0,01385883172016	0,01385883172016
3	в	43	0,0166731	0,18766963	3	в	40	0,0133550	0,21569282	3	в	36	0,0112710	0,199749	0,015943290999505	0,015943290999505
4	г	40	0,0155098	0,20317952	4	г	46	0,0153580	0,23105175	4	г	62	0,0194110	0,219160	0,011890826183903	0,011890826183903
5	ф	1	0,0003877	0,20356727	5	ф	0	0	0,23105175	5	ф	40	0,0125230	0,231684	0,00063265534396198	0,00063265534396198
6	д	169	0,0655292	0,26909654	6	д	173	0,0577620	0,28881469	6	д	174	0,0544770	0,286161	0,0026531382402109	0,0026531382402109

Рисунок 3.3.2 – Вычисления ЦП униграммы

В подглаве 3.4 представлены основные выводы третьей главы. В заключении диссертационной работы приведены итоговые результаты исследования, оформлены основные выводы и определены важные направления будущих исследований, связанных с темой данной диссертации.

ЗАКЛЮЧЕНИЕ

Ниже перечислены основные результаты диссертационной работы:

- ✓ анализ вопроса благозвучия текста и путей его решения в доступных источниках и разработка основ эвфонометрии словоформ естественных языков, в том числе таджикского языка [1-А, 2-А, 11-А];
- ✓ компьютерная обработка текста произведения Саади «Гулистан», выявляющая статистические данные и особенности сочетания звуков в его словесном составе [3-А];
- ✓ отобрана большая коллекция текстов и ее элементы: униграмма, биграмма и триграмма исследованы с математической и компьютерной точки зрения [4-А, 5-М, 6-А];
- ✓ нахождение труднопроизносимых признаков на основе опроса носителей языка, классификации таджикских словоформ и разработки компьютерных программ для описания символов [7-А, 8-А];
- ✓ Исследованы таджикские слоги, разработаны алгоритм и программа разделения слогов таджикских слов [10-А, 12-А];
- ✓ для решения задачи автоматического распознавания благозвучие словоформ таджикского языка создан программный комплекс на языке программирования Delphi [13-А].

Рекомендации по практическому использованию результатов.

Результаты исследования могут быть использованы при решении задач по языкознанию и литературоведению, проведении различных речевых тестов по конкретным вопросам, фонетическим системам и учебно-контрольным системам. Результаты компьютерных расчетов показывают, что определение их качественных и количественных характеристик имеет большое практическое значение для классификации словоформ на классы. Сообщается,

что результаты исследования диссертационной работы для чтения спецкурсов («Методы анализа текста на компьютере», «Лингвистические основы машинного перевода», «Компьютерные технологии в лингвистических исследованиях»), выполнения лабораторных и самостоятельных работ, курсовых, дипломных и магистерских работ для студентов и магистров по направлениям «Компьютерная лингвистика», «Лингвистика и новые информационные технологии» и «Информатика» в вузах Республики Таджикистан.

Комплекс компьютерных программ, созданных в рамках данной диссертации, применяется в сферах науки и образования, культуры, филологии, компьютерных систем обработки текста, машинного обучения и систем искусственного интеллекта для решения важные задачи компьютерной лингвистики.

Публикации по теме диссертации
Статьи, опубликованные в научных журналах, рекомендованных ВАК
при Президенте Республики Таджикистан

- [1 – А] Пиров С.М. К разработке основ эвфонометрии слов [Текст] / З.Д. Усманов, С.М. Пиров // Известия Академии наук Республики Таджикистан. Отделение физико-математических, химических, геологических и технических наук. - 2015. - №3 (160). - С. 39-42.
- [2 – А] Пиров С.М. Дар бораи баъзе масъалаҳои хушоҳангии матнҳои тоҷикӣ [Матн] / С.М. Пиров // Паёми Донишгоҳи миллии Тоҷикистон. -2017, -№ 4-3. -С.105-109.
- [3 – А] Пиров С.М. Таҳқиқи биграммаҳои “Гулистон”-и Саъдии Шерозӣ [Матн] / С.М. Пиров // Паёми Донишгоҳи миллии Тоҷикистон. -2017, -№ 1/5. - С.156-159.
- [4 – А] Пиров С.М. Алгоритм распознавания благозвучия таджикских слов при использовании униграмм [Текст] / С.М. Пиров // Доклады Национальной академии наук Таджикистана. -2022, т. 65, № 1-2. С.57-62.
- [5 – А] Пиров С.М. Алгоритми шинохти хушоҳангии калимаҳои тоҷикӣ бо истифодаи биграммаҳо [Матн] / С.М. Пиров // Паёми ДМТ, бахши илмҳои табиӣ. -2022. -№2. - С. 102-111.
- [6 – А] Пиров С.М. Алгоритм распознавания благозвучия таджикских слов при использовании триграмм [Текст] / С.М. Пиров // Вестник, Филиала МГУ им. В.М.Ломоносов в г.Душанбе, раздел информатика. -2023, Том 1, -№1 (29). -С. 35 - 42.

Статьи в материалах научных конференций

- [7 – А] Пиров С.М. Дар бораи гурӯҳбандии калимаҳои забони тоҷикӣ аз рӯйи талаффузашон [Матн] / С.М. Пиров // РТСУ. Перспективы развития фундаментальных и прикладных лингвистических исследований в Республике Таджикистан. – Душанбе: Персис, 2016. - 1(7). -С. 16 -18.
- [8 – А] Пиров С.М. Дар бораи маҷмӯи аломатҳои тавсифкунандаи хушоҳангии калимаҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи конференсияи илмӣ-амалии байналмилалӣ. ДТТ. -Душанбе -2017. -С. 352-355.
- [9 – А] Пиров С.М. Басомади биграммаҳои ҳарфҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи Конференсияи ҷумҳуриявӣ илмӣю назариявӣ ҳайати устодону кормандони ДМТ баҳшида ба ҷашнҳои «5500-солагии Саразми бостонӣ», «700-солагии шоири барҷастаи тоҷик Камоли Хучандӣ» ва «20-солаи омӯзиш ва рушди фанҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илму маориф (солҳои 2020-2040)». – Ҷилди 1. -2020. -С. 39-40.
- [10 – А] Пиров С.М. Алгоритми содаи ба ҳичоҳо ҷудокунии калимаҳои тоҷикӣ [Матн] / С.М. Пиров // Маводи конференсияи байналмилалӣ илмӣ-амалӣ дар мавзӯи «Муаммоҳои муосири математика ва татбиқи он» баҳшида ба «Солҳои

2020-2040 эълон гардидани 20-солаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ». -2022. -С. 158-161.

[11 – А] Пиров С.М. Таҳлили моделҳои баҳодиҳии хушоҳангии матнҳо [Матн] / С.Х. Мирзоев, С.М. Пиров // Маводи конференсияи байналмилалӣ илмӣ-амалӣ дар мавзӯи «Таҳлили компютерӣ масъалаҳои илм ва технология» бахшида ба «Солҳои 2020-2040 эълон гардидани 20-солаи омӯзиш ва рушди илмҳои табиатшиносӣ, дақиқ ва риёзӣ дар соҳаи илму маориф» ва «75-солагии Донишгоҳи миллии Тоҷикистон». -2023. -С. 396-404.

Авторские права и свидетельство:

[12-А]. Пиров С.М. Свидетельство. Чудокунӣ ҳичоҳи калимаҳои тоҷикӣ / С.М. Пиров // Свидетельство об официальной регистрации программы для ЭВМ в отделе по правам и защите прав автора Министерства культуры Республики Таджикистан (№ 150, 9.11.2022г.).

[13-А]. Пиров С.М. Автоматизированная система анализа текстов и оценка благозвучия таджикских слов [Текст] / С.Х. Мирзоев, А.А. Косимов, **С.М. Пиров**// Свидетельство о государственной регистрации информационного ресурса №1202500527 от 28.02.2025 внесен в реестр информационных ресурсов Республика Таджикистан.

Список литературы

[1]. Мудрые слова Президента Эмомали Рахмона о языке. – Душанбе: Бунёди забон, 2015. – 116 стр.

[2]. Усманов З.Д. Классификатор дискретных случайных величин –Доклады Академии наук Республики Таджикистан, 2017, т.60, № 7-8. – С. 291-300.

[3]. Усманов З.Д. Алгоритм настройки кластеризатора дискретных случайных величин // Доклады АН РТ, 2017, т.60, № 9, – С. 392-397.

ШАРҲИ МУХТАСАР

рисолаи Пиров Саидали Мирзоалиевич дар мавзуи «Таҳия ва тадқиқи критерияҳои миқдорӣ барои баҳодиҳии хушоҳангии калимаҳои тоҷикӣ», барои дарёфти дараҷаи илмӣ номзади илмҳои техникӣ аз рӯйи ихтисоси 05.13.11 – «Таъминоти математикӣ ва барномавии мошинҳои ҳисоббарор, комплексҳо ва шабакаҳои компютерӣ»

Калимаҳои калидӣ: дастгоҳи гуфтор, овоз, алифбо, забони тоҷикӣ калима, аломат, хушоҳангӣ, душворталаффузӣ, басомад, бузургҳои тасодуфӣ, симои рақамӣ, алгоритмҳои коркарди матн, таснифгар, баҳодиҳӣ, пайкараи матнҳо, таъминоти барномавӣ, амсилаҳои баҳодиҳии хушоҳангӣ.

Объекти таҳқиқот – амсилаҳои математикию компютери таҳлил ва баҳодиҳии хушоҳангии маҷмуи калимашаклҳои ҷопии забони тоҷикӣ.

Мавзуи таҳқиқот – шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ дар асоси аломатҳо ва басомади N -граммаҳои он.

Усулҳои таҳқиқот. Барои ҳалли мақсаду вазифаҳои дар тадқиқот гузошташуда аз методҳои фонетикӣ, таҳлили омӯрӣ, ҳисобҳои таҷрибавӣ, назарияи маҷмуъҳо, таҳлили системавӣ, амсиласозии математикӣ ва компютерӣ, барномасозии ба объект нигаронидашуда барои таҳияи воситаҳои барномавӣ ва коркарди манбаи додаҳои релятсионӣ истифода шудааст.

Навгониҳои илмӣ тадқиқот.

- иттилоъдиҳии аломатҳое, ки барои шинохт ва муайянкунии хушоҳангии калимашаклҳо таъйин шудаанд таҳқиқ карда шуд;

- самаранокии истифодаи N -граммаҳои ҳуруфӣ ($N=1,2,3$) ба сифати аломатҳои муайянкунандаи хушоҳангии калимашаклҳо муқаррар карда шуд;

- самаранокии таснифгари Усмонов З.Ҷ. муқаррар карда шуд, ки қобилияти бо саҳеҳии то 96% муайян кардани хушоҳангии маҷмуи калимашаклҳоро дорад;

- бори аввал дар Ҷумҳурии Тоҷикистон маҷмааи барномаи компютери замонавӣ барои шинохти худкори хушоҳангии калимашаклҳои тоҷикӣ, сохта шуд.

Дараҷаи эътимоднокии натиҷаҳои рисола. Натиҷаҳои бадастомада, бо як силсила озмоишҳои ҳисоббарорӣ тасдиқ мешаванд, ки дар онҳо хушоҳангии калимашаклҳои гуногун бо саҳеҳияти кифоя баланд муайян шудаанд.

Соҳаи истифодабарӣ: Натиҷаҳои бадастомадаро дар соҳаҳои илму маориф, фарҳанг, забоншиносӣ, адабиётшиносӣ, низомҳои компютери коркарди матн, зехни сунъӣ ва омӯзиши мошинӣ барои ҳалли масъалаҳои мубрами забоншиносии компютерӣ истифода бурдан мумкин аст.

РЕЗЮМЕ

диссертация Пиров Саидали Мирзоалиевич на тему «Разработка и исследование количественных критериев для оценки благозвучия таджикских слов», на соискание ученой степени кандидат технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Ключевые слова: речевой аппарат, звук, алфавит, таджикский язык, слово, признак, благозвучие, труднопроизносимые, частота, случайные величина, цифровой портрет, алгоритмы обработки текста, классификатор, оценка, корпус текстов, программное обеспечение, модели оценки благозвучие.

Объект исследования. В качестве объекта исследования определены математические и компьютерные модели анализа и оценки множества печатных словоформ таджикского языка.

Предмет исследования. Автоматическое распознавание благозвучия таджикских словоформ на основе признаков и частотности его N -грамм.

Методы исследования. Для решения целей и задач, поставленных в исследовании, используются фонетические методы, методы математической статистики, вычислительного эксперимента, теории множеств, системного анализа, распознавания, математические и компьютерные моделирование и объектно-ориентированного программирования для разработки программных средств.

Научная новизна работы. Основные результаты, приведенные в диссертации, являются новыми, и они следующие:

- исследована информативность признаков, предназначенных для распознавания благозвучия словоформ;
- установлена эффективность использования N -граммы букв ($N=1, 2, 3$) в качестве признаков, определяющих благозвучия словоформ;
- показана эффективность применения классификатора Усманова З.Дж., который способен распознавать благозвучие коллекции словоформ с точностью до 96 %;
- впервые в Республике Таджикистане создан компьютерный программный комплекс для распознавания благозвучия таджикских слов.

Степень достоверности результатов диссертации. Полученные результаты подтверждаются серией вычислительных экспериментов по идентификации с достаточно высокой точностью в различных коллекциях текстов.

Область применения: Полученные результаты могут быть использованы в сферах науки и образования, культуры, филологии, компьютерных систем обработки текста, систем искусственного интеллекта и машинного обучения для решения актуальных задач компьютерной лингвистики.

SUMMARY

of the thesis of Pirov Saidali Mirzoalievich on the topic "Development and research of quantitative criteria for assessing the euphony of Tajik words", submitted for the degree of candidate of technical sciences by specialty 05.13.11 – “Mathematical and software computing machines, complexes and computer networks”

Key words: speech apparatus, sound, alphabet, Tajiki language, word, feature, euphony, difficult to pronounce, frequency, random variable, digital portrait, text processing algorithms, classifier, assessment, text corpus, software, euphony assessment models.

Object of the research. The object of the research is the definition of mathematical and computer models of analysis and evaluation of the set of printed word forms of Tajik language.

Subject of the research. Automatic recognition of euphony of Tajik words based on features and frequency of its N-grams.

Research methods: Phonetic methods, methods of mathematical statistics, computational experiment, set theory, system analysis, mathematical and computer modeling, object-oriented programming for the development of software and relational database processing are used for solution the goals and objectives set in the study.

The scientific novelty of the work.

- the informativeness of the signs assigned to the recognition and determination of the harmony of Tajiki words was investigated;
- the efficiency of using letter N-grams ($N=1, 2, 3$) as signs determining the harmony of word forms was established;
- the efficiency of the modifier of Usmanov Z.D., capable of recognizing the euphony of words with an accuracy of up to 96%, was established;
- for the first time in the Republic of Tajikistan, a modern computer program complex for automatic recognition of the euphony of Tajiki words was created.

The degree of reliability of thesis results.

The obtained results are confirmed by a series of computational experiments to identify authors with sufficiently high accuracy in various text collections.

Field of the application. The results of the thesis can be used in the fields of science and education, culture, linguistics, philology, computer systems for processing text, artificial intelligence systems and machine learning to solve urgent problems of computer linguistics.

Сдано в печать 30.05.2025 Формат 60x84
1/16, тираж. 60 экзем.
Отпечатано в типографии ТНУ
г. Душанбе.