

ТҮРКОЛОГИЯДА ЛИНГВОСТАТИСТИКАНЫН ПАЙДА БОЛУШУ ЖАНА КАЛЫПТАНЫШЫ

Илимий макалада түрк тилдеринде лингвостатистиканын башталышы жана түркологиядагы мындай ыкма менен жүргүзүлгөн илимий изилдөөлөр тарыхый жактан иликтенет.

Электрондук техника дүйнө жүзүнө тарап, азыркы учурда өтө тездик менен өнүгүп жаткан мезгил. Электрондук техника инженердик-экономикалык милдеттерди гана чечпестен, ал тил илими боюнча да негизги маселелерди чечүүгө өз жардамын тийгизип жатат.

Мисалы, электрондук-эсептөө машиналарынын жардамы менен бир тилден экинчи тилге которууга, илимий-техникалык тексттердин рефераттары менен аннотацияларын автоматтык жол менен жасоого, керектүү информацияларды жеңил таап алууга, оозеки речти автоматтык түрдө анализдөө жана синтездөөгө кеңири жол ачууда. Бул алгоритм жолу боюнча ишке ашат.

Эмгекти көп талап кылуучу тил илиминин практикалык иштерин автоматташтыруу жолу менен бөлүштүрүү кеңири пайдаланылып жатат. Индоевропалык тил илиминин адистери колдонгон математикалык жана кибернетикалык методдордун тажрыйбасы түрк таануу багытындагы изилдөөлөргө да таасирин тийгизбей койгон жок.

Өткөн кылымдын аягында индоевропалык тил таануу илимине кирген статистикалык ыкма түрк тилдерине 60-жылдын башталышында киргизилген. Бул болсо түрк тилдеринин агглютинативдик түзүлүштөгү тил экендиги менен түшүндүрүлөт.

Академик А.Н. Кононов: Петр Инин орус илиминин өнүгүшүнө киргизген эң жогорку салымы, бул илимдер академиясын (1724-1725) түзгөндүгү болуп саналат

деген [13]. Анда математика жана физика класстары менен бирге эле гуманитардык класстар да болгон. Математика жана филология сабактары өз ара тыгыз байланышта жана тең укукта өнүгө баштаган. Биринчи орус академиги М.В. Ломоносов да академиялык окуу китебинин грамматикасын, орус ырларынын түзүлүш теориясын жана орус адабият тилинин түзүлүш принциптерин иштеп чыккан.

Андан кийин П.Л. Чебышев (1821-1894), М.В. Остроградский (1801-1861), В.Я. Буняковский (1804-1889) деген окумуштуулар математика илими менен бирге тил илиминин маанилүү ролун танган эмес.

Эгерде В.Я. Буняковский биринчи жолу статистиканы тил илимине колдонуу мүмкүндүгү жөнүндө өз оюн билдирсе, башка бир атактуу математик А.А.Марков 1913-жылы биринчи жолу статистиканы тилдик материалга колдонгон. Булардан кийин орус окумуштуулары В.А.Богородицкий, А.М.Пешковский чыгармадагы тыбыштарга статистикалык анализ жасаган.

Жыйырманчы жылдын аягында, отузунчу жылдын баш ченинде тилди математикалык жол менен изилдөө иштери (бирдей көз карашта болбосо да), Орто Азия жана Казакстанга тарай баштаган.

Ошентип, латын алфавитине өтүүгө даярдык көрүп жаткан учурда профессор Н.Ф.Яковлев өзүнүн математикалык формуласын алфавиттик тартипте жыйынтыктап чыгат [17]. Бул окумуштуунун аты Москвада 1968-жылы чыккан Е.Д.Поливановдун бир томдук жыйнагында үч жолу жолугат. Бирок анын математикалык жол менен изилдөө иштерине көңүл деле бурулган эмес.

Бардык эле илимдин тармагы боюнча орус тилиндеги жана чет өлкөдөгү илимий жетишкендиктер акырындап башка тилдерге, б.а. түрк тилдерине да таасирин тийгизе баштаган. Ошонун сыңарындай эле лингвистикадагы статистикалык изилдөөлөрдүн башталышын, албетте, орус илимпоздорунун эмгектеринен кезиктиребиз.

Акыркы жылдары математикалык метод тил илиминде эле активдүү пайдаланылбастан, социологияда, психологияда, археологияда ж.б. илимдин тармактарында кеңири колдонула баштады. Тил илиминде ар кайсы математикалык методдордун ичинен өзгөчө жакшы текшерилип, көп колдонулуп, сыноодон өткөн жана азыр дайыма пайдаланылып жүргөн статистикалык-ыктымал методу (вероятностно-статистический метод) болуп саналат. Бул метод тил боюнча мүмкүн деп элестете албаган закон-ченемдүүлүктөрүн текшерип бере алат. Сыноодон өткөн бул метод түрк тилдерин, анын ичинде казак, өзбек, азербайжан, кыргыз, каракалпак тилдерин изилдөө иштеринде ийгиликтүү колдонулуп келген жана мындан ары да колдонула бермекчи.

1968-жылы Казак илимдер академиясында тил илими институту түзүлгөндөн кийин лингвистикалык-статистикалык изилдөө иштери системалуу жана натыйжалуу түрдө жүргүзүлө баштады. Мында атайы "Автоматизация лингвистических работ и статистика речи" деген атайын группа түзүлгөн. Казак лингвостатисттери 20 жылдын ичинде 10 монография жана изилдөө жыйнактарын, 200дөн ашуун илимий жана илимий-популярдуу макалаларды чыгарган. 1973-жылы 30 басма табактагы "Статистика казахского текста" деген илимий макалалардын жыйнагы чыккан. Бул 10 жыл ичинде иштелген иштин жыйынтыгы болуп саналат. "Статистика казахского текста"-деген жыйнак коллективдүү иштелген 1-жыйнактын уландысы болуп кетет. Азыркы учурда ЭЭМдин жардамы менен ар кандай көлөмдөгү жыштык сөздүктөр,

М.О.Ауэзовдун 20 томдук чыгармалар жыйнагы боюнча автоматташтырылган картотекалык фонд түзүлгөн. Булар орус жана казак тилинде чыгат.

Мындан башка түрк тилдери боюнча этапташтырылган машиналык фонд түзүү процесси иш жүзүнө ашырыла баштаган. Бул методду 1-жолу казак окумуштуулары колдонушкан. Ушул максатта Казак илимдер академиясынын тил институтунда атайын түзүлгөн лингвистикалык-статистикалык изилдөө жана автоматташтыруу тобу бүткүл союздук "Статистика речи" тобунун ажырагыс бир бөлүгү катарында иш жүргүзүп келген. Азыр Казакстанда математика-лингвистикалык эки топ Алматы жана Чымкент шаарларында иштейт. Алар казак тексттери боюнча лексикографиялык иштерди автоматташтыруу жана статистикалык изилдөө иштерин жүргүзөт. Ошондой эле машинада которуу проблемасы менен иштешет. Бул иштин жалпы жыйынтыгы 1973-жылы чыккан "Статистика казахского текста" деген жыйнакта чагылдырылган [14]. Мына ушуну менен казак окумуштуулары математикалык лингвистиканы изилдөөнүн атайын мектебин түздү.

Казак окумуштуулары ушул жылдардын аралыгында газета, жомок, көркөм адабий чыгармалардын алфавиттик жыштык сөздүгүн, тексттердин графем-фонемалык жыштыгын, сөздүн маани билдирүүсүн өлчөө, казак тилинин энтропиясын изилдөө иштерин жүргүзгөн. Биринчи жолу казак эле эмес, жалпы түркологиялык тексттерди автоматташтыруу жана моделдештирүү проблемасын козгогон. Авторлор Орхон-Енисей жазуу эстеликтеринин белгилери үчүн машиналык индексти сунуш кылышкан.

Эсептөөчү техниканы пайдалануу менен тилди үйрөнүүнүн ыктымалдык-статистикалык жана автоматташтыруу процесси менен бирге К.Б. Бектаев иштеп чыккан. Анын "Математические методы в языкознании" деген монографиясы бардык лингвисттер үчүн (ушул методду колдогондор үчүн) фундаменталдуу курал болуп саналат. [3]

К.Б. Бектаевдин "Статистико-информационная типология тюркского текста" деген монографиясында биринчи жолу статистикалык-информациялык типологиянын проблемалары каралган [4].

Автор бул эмгегинде индоевропалык тил менен түрк тилдерин салыштырып, андагы типологиялык өзгөчөлүктөрдү тапкан. Алматыда 1969-жылы "Статистическое и информационное изучение тюркских языков" деген бүткүл союздук семинар өткөрүлгөн. Анда негизги проблемалар каралган эле. Казак Илимдер Академиясынын тил институтунда ЭЭМдин жардамы менен индекстөө жана колдонуунун жалпы принциптерин автоматташтырып түрк тексттерин түзүүнү К.Б. Бектаев жана А.Х.Жубановдор иштеп чыгышкан.

Түркологияда "Абай тилинин сөздүгү" деген жыштык сөздүк түзүлүшүн алгачкы саамалык болгон [1]. Ал традициялык метод менен эле иштелип, сөздөрдүн колдонуш жыштыгын аныктаган түшүндүрмө сөздүк болуп эсептелет.

Бүгүнкү күндө казак, өзбек тилдеринде айрым авторлордун чыгармаларынын, публицистиканын тилин, балдар адабияты менен илимдин тилин толук кучагына камтыган жыштык сөздүктөр түзүлдү.

Элге кеңири белгилүү болгон орто кылымдагы жазуу эстелик "Кодекс Кулсаникус" өз учурунда латын, немец, француз тилдерине которулган. Бирок СССР элдеринин тилдерине которулган эмес. Бул эстеликтин казак тилинде чыгышы жана жыштык көрсөткүчтөрүнүн берилиши жалпы эле түрк тилдери үчүн өтө чоң жетишкендик болуп саналат.

М.О.Ауэзовдун "Абай жолу" романы боюнча (1979-жылы Алма-Атада чыккан) түзүлгөн 1-жыштык сөздүк улантылын, кийин 20-томдук М.О.Ауэзовдун чыгармаларынын жыйнагы боюнча изилдөөлөр жүргүзүлгөн.

Бул лексикографиялык эмгекте жазуучунун лексикалык жана фразеологиялык сөз байлыгын, сөздү пайдалануудагы жекече стилдик өзгөчөлүгүн жана казак тилинин көркөм сүрөттөөчү каражаттарын кантип чагылдырып пайдаланганын табат. Ал жазуучунун тилдик өзгөчөлүгүн үйрөнүү үчүн, ошондой эле жалпы казак лексикасынын түзүлүшүн статистикалык жактан изилдөө үчүн өтө бай материалды берет.

М.О.Ауэзовдун чыгармаларынын текстин изилдегенде, анын көлөмү 2,5 миллиондон ашуун сөздү түзгөн. Бул ишти улантууну пландаштырган. Ошол эле тил илими институтунда каракалпак жана кыргыз тилиндеги жыштык сөздүктөр түзүлүп бүтүп, аяктоо мезгилинде турат. Публицистиканын тили изилденип жатат.

Кыргыз тили боюнча биринчи болуп Кыргыз мамлекеттик университетинин профессору К. Дыйканов тарабынан жөнөкөй саноо жолу менен "Манас" эпосунун бириктирилген вариантынын 1-жолу боюнча эки бөлүктөн турган жыштык сөздүгү түзүлгөн [11]. Бирок бул сөздүк "Манас" эпосунун бай лексикасын толук көлөмдө чагылдыра албайт. Ошондой эле сөздүк бириктирилген варианттын негизинде түзүлгөндүктөн, ар башка аткаруучунун сөз байлыгын, сүрөттөө каражаттарынын өзгөчөлүгүн да бере алган эмес. Ошондуктан аны жакшы иштин башталмасы катары гана баалого болот.

Доцент Ж. Мамытовдун жалпы илимий жетекчилиги алдында Ысык-Көл мамлекеттик университетинин тилчилер жана информатиктер тобу ЭЭМди (электрондук-эсептөө машинасын) колдонуу менен заманыбыздын залкар манасчылары Сагымбай Орозбаковдун 4 томдук "Манасынын" жана Саякбай Каралаевдин 5 томдук ("Манас", "Семетей", "Сейтек") трилогиясынын тексти боюнча түзүшкөн, болжолу 165 басма табак көлөмүндөгү "Манас" эпосунун лексикасынын жыштык сөздүгү өзүнүн басылышын күтүп жатат. Биз өзүбздүн изилдөө ишибизде ушул сөздүктүн материалдарынан да пайдаланмакчыбыз. Айрым авторлордун чыгармалары боюнча жыштык сөздүктөрдү, бардык адабий тилдин функционалдуу стилин түзүү иши абдан актуалдуу маселе болуп саналат. Булардын негизинде тилдин лексико-грамматикалык нормасын аныктоого мүмкүн. Бул сөздүктөрдү түрк тилдеринин салыштырма-типологиясын үйрөнүүдө, ошондой эле түрк тилдеринин машиналык фондусун түзүүдө пайдаланса болот.

Ошол эле Алматыда коргогон А.Ахабаевдин тилди нормативдик планда үйрөнүү боюнча жазылган диссертациясы бир топ кызыгууларды туудурган. Анда азыркы кездеги публицистиканын тилин лексико-морфологиялык түзүлүшү боюнча изилдөөгө арналган. [2]

Статистика-морфологиялык анализдин жардамы менен казак тилиндеги грамматикалык формалардын варианттарын стилдик жактан дифференциялоо маселеси изилденген. Атап айтсак, А.Х. Жубанов, М. Ауэзовдун "Абай жолу" романындагы сын атоочторду лексика-морфологиялык каражаттарын аныктап чыккан болсо, С.Мырзабеков азыркы казак тилиндеги этиштин структуралык түзүлүшүн изилдеген [10].

Каракалпак тилиндеги ат атоочтор боюнча авторлор коллективи анализ жүргүзгөн. Ошону менен катар А. Бабанаров азыркы түрк тилиндеги газетанын жыштык сөздүгүн түзгөн [5].

Ж. Жетекишов кыргыз тилинде чыккан публицистикалык чыгармалардагы зат атоочтун сөз өзгөртүүчү мүчөлөрү боюнча статистикалык анализ жүргүзүп, аны казак тилинин материалы менен салыштырып иштеген.

Ал эми азербайжан тилиндеги сөз жасоочу мүчөлөрдүн сандык мүнөздөмөсүн С.Мамедов изилдеген [12].

Булардан өзгөчөлөнүп, өз ара семантикалык жактан изилдөө үчүн салыштырып Т.М. Гарипов статистикалык анализден пайдаланган. Натыйжада башкыр, казак, татар, чуваш ж.б. тилдери сандык жана сапат жагынан бири-биринен кескин айырмалана тургандыгы аныкталган. [8]

Жогоруда аталгандардан башка да ар кайсы тилдердин грамматикалык формаларын статистикалык жактан салыштырып изилдеген изилдөөлөр бар. Түрк тилдеринин тыбыштык түзүлүшүн эски традициялык (угуу) методу аркылуу текшерип изилдөө менен катар инструменталдык эксперименталдык жолдор менен статистикалык изилдөө иштери да жүргүзүлүүдө.

Бул багыттагы иштердин катарына кумандылыктардын тили боюнча И.Я.Селотинанын, азербайжан тили боюнча Ф.С. Вейсалов, Я.К. Керимов, М.Т.Сафаровдордун, татар тили боюнча Т.И. Ибрагимовдун, өзбек тили боюнча Д.А.Байтанаеванын жана башкалардын изилдөөлөрүнүн натыйжасында түрк тилдериндеги фонемалардын колдонуш жыйынтыгы, алардын варианттары, айкашуу касиеттери, муундун структуралык типтери, тутуму, ачык жана жабык түрлөрү, сөздөгү муундун орду жана абалы жөнүндө мүнөздөмө берилип, фонемаларга статистикалык-информациялык өлчөөлөр жүргүзүлгөн [15].

Фонетиканы статистикалык метод менен изилдөөнүн натыйжасында тилдин ар түрдүү жанрларында информациялык мүнөздөмөлөр ар башкача болорлугун аныктоого мүмкүнчүлүк түзүлүп отурат. Маселен, казак тилинин сүйлөө жана ырдоо речине статистикалык мүнөздөмө берилген [7].

Д.А.Байтанаева кийинки эмгектеринде эксперименттик тажрыйбанын негизинде казактын жазма тексттериндеги муундун статистикалык-информациялык касиетин изилдеген. Статистикалык метод түрк тилдеринин синтаксистик түзүлүшүн изилдөөдө да ийгиликтүү колдонулууда. [6]

М.Р.Жумабаева англис жана казак тили боюнча жөнөкөй сүйлөмдөгү сөздөрдүн орун тартибинин өзгөчөлүгүнө карата салыштырма-статистикалык анализ жасаган [9].

А.Х. Жубановдун "Квантативная структура казахского текста" (ЭЭМде лингвистикалык анализдин тажрыйбасынан) деген монографиясында (Жубанов А.Х. Квантативная структура казахского текста. Алма-Ата, 1987.с.147.) түрк тилдериндеги тексттердин лингвистикалык жыштык тизмесин автоматташтыруу тажрыйбасы, ошондой эле, сөздүк-сөз көрсөткүчтөрдү алуунун алгоритми каралган. Казак тилиндеги сөздөрдү ыктымалдык-статистикалык үлгүдө классификациялоо автор тарабынан өзгөчө баса белгиленип, лингвистикалык статистиканын ар кандай маселелерине токтолгон

Кыргыз тилинде бул багытта бир кыйла изилдөө иштери жүргүзүлүп, диссертациялар корголду. Алардын ичинен эң негизги айрым иштерге токтолуп көрөлү.

С.И. Ибрагимов кыргыз тилиндеги зат атоочторду синтаксистик-функционалдык жактан классификациялап, зат атоочтун башка сөз түркүмдөрү менен болгон синтаксистик байланышын аныктаган. Ошону менен бирге, кыргыз тилинде сөздөрдүн орун тартибин изилдөөдө статистикалык анализди колдонгон.

Натыйжада, илимий жана көркөм адабияттардагы жөнөкөй сүйлөмдөрдүн түзүлүш өзгөчөлүгүн аныктоого мүмкүнчүлүк түзүлдү [10].

Т.Садыковдун "Түрк морфологиясын моделдештирүү проблемасы" деген эмгеги кыргыз тилиндеги атоочтук сөз формаларына арналган [16].

Эмгекте кыргыз тилинин морфологиялык структурасы (түзүлүшүн) анализделип, алынган анализдин жыйынтыгы лексико-грамматикалык информацияга берилген атоочтук сөз формаларынын моделин (үлгүсүн) түзүү максатында пайдаланылат.

Жогоруда ыктымалдык-статистикалык методду колдонуу менен түрк тилдерин изилдеген эмгектерге токтолдук.

Тилдин бардык деңгээлдерин изилдөөдө бул метод эң кеңири колдонулганын жана универсалдуу (көп тармактуу) экенин көрдүк.

Казак, кыргыз, азербайджан, өзбек тилдеринде фундаменталдуу эмгектер жаралып жатканына, ошондой эле айрым эмгектер али да жогорку деңгээлге жете элегине күбө болдук. Бул метод мындан ары өркүндөтүлүп, илимдин бардык тармактарында ийгиликке жетерине ишенебиз.

АДАБИЯТТАР

1. Абай тілі сөздігі. -Алматы, 1968.
2. Ахабаев А. Статистический анализ лексико-морфологической структуры языка казахской публицистики. -Алма-Ата, 1971.
3. Бектаев К.Б., Пиотровский Р.Т. Математические методы в языкознании. Ч.І.: Теория вероятности и моделирование нормы языка. -Алма-Ата, 1973. -С.282; Бектаев К.Б., Пиотровский Р.Т. Математические методы в языкознании. Ч.ІІ.: Математическая статистика и моделирование текста. -Алма-Ата, 1974. -С.336.
4. Бектаев К.Б. Статистико-информационная типология тюркского текста. -Алма-Ата, 1978. -С.183.
5. Бабанаров А. Частотный словарь и автоматический словарь для машинного перевода турецких газетных текстов. -Л., 1980.
6. Байтанаева Д.А. Информационные характеристики казахского текста. -Алма-Ата, 1985. -С.18.
7. Вейсалов Ф.Е., Керимов Л.К., Сафаров М.Т. Комбинаторика фонем современного азербайджанского языка. -Баку, 1981.
8. Тарипов Т.М. Лексикостатистические отношения тюркских языков Урала и Поволжья. -Уфа, 1972.
9. Жумабаева М.Р. О статистической характеристике простого предложения в английском и казахском языках. -Алма-Ата, 1980. -С.62-63.
10. Жубанов А.Х. Статистический анализ лексико-морфологической структуры языка казахской публицистики. -Алма-Ата, 1971. -С.24.
11. Ибрагимов С.И. Некоторые характеристики имен существительных киргизского языка //Тезисы докладов и сообщений всесоюзного семинара. -Алма-Ата, 1969. -С.34-38.

Филологические науки

12. "Манас" эпосунун алфавиттүү сөздүгү: I бөлүк. –Фрунзе, 1988. "Манас" эпосунун жиги алфавиттүү лугаты: II бөлүк. –Фрунзе, 1988.
13. Мамедов Э.С. Количественные характеристики словообразовательных элементов. –Баку, 1979. -С.105-110.
14. Кононов А.Н. Математические методы в историко-экономических и историко-культурных исследованиях. –М., 1977.
15. Статистика казахского текста. –Алма-Ата, 1973.
16. Селютина И.Я. Относительная частотность согласных в языке кумандинцев. – Новосибирск, 1981.
17. Садыков Т. Проблемы моделирования тюркской морфологии. –Фрунзе: 1987.- С.121.
18. Яковлев Н.Ф. Математическая формула построения алфавита //Культура и письменность Востока. -1928. -№1.