

P.2008  
10014

Асқар ЖҰБАНОВ

ҚОЛДАНБАЛЫ  
ТІЛ БІЛІМІНІҢ  
МӘСЕЛЕЛЕРІ



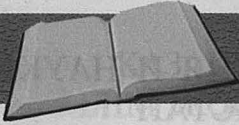
ВОПРОСЫ  
ПРИКЛАДНОГО  
ЯЗЫКОЗНАНИЯ



*Mydas*



Р2008/1001К



# Тіл білімінің озық үлгілері

АҚПАР ЖҰБАНОВ

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ  
БІЛІМ АЛҒАШҚЫ ҚАЖАҚТАУ



**АРЫС**

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РЕСПУБЛИКИ КАЗАХСТАН  
ИНСТИТУТ ЯЗЫКОЗНАНИЯ им. А.БАЙТУРСЫНОВА

МИНИСТЕРСТВО КУЛЬТУРЫ И ИНФОРМАЦИИ  
РЕСПУБЛИКИ КАЗАХСТАН  
КОМИТЕТ ПО ЯЗЫКАМ

**Аскар ЖУБАНОВ**

**ВОПРОСЫ ПРИКЛАДНОГО  
ЯЗЫКОЗНАНИЯ**

Издательство "Арыс"  
Алматы  
2008

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ  
А.БАЙТҰРСЫНҰЛЫ АТЫНДАҒЫ ТІЛ БІЛІМІ ИНСТИТУТЫ

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
МӘДЕНИЕТ ЖӘНЕ АҚПАРАТ МИНИСТРЛІГІ  
ТІЛ КОМИТЕТІ

**Асқар ЖҰБАНОВ**

**ҚОЛДАНБАЛЫ ТІЛ БІЛІМІНІҢ  
МӘСЕЛЕЛЕРІ**

“Арыс” баспасы  
Алматы  
2008

ББК 81.2-5

Ж 80

*Кітап “Арыс” қорының өз қаражатымен “Қазақ ғылымының озық үлгілері” атты мегажобасының аясында жарық көріп отыр.*

*Книга вытущена фондом “Арыс” на собственные средства в рамках мегапроекта “Лучшие образцы казахской науки”.*

**Ж 80 ЖҰБАНОВ А.**

Қолданбалы тіл білімінің мәселелері. Вопросы прикладного языкознания. – Алматы: “Арыс” баспасы, 2008. – 640 бет. + 8 бет суретті жапсырма.

ISBN 9965-17-460-1

Қолданбалы қазақ тіл білімінде зор еңбек етіп келе жатқан тілші-математик Асқар Жұбановтың зерттеулер жинағында статоллингвистика, математикалық және компьютерлік лингвистика салаларының әдіс-тәсілдері мен қыр-сырлары, бүгіні мен болашағы жан-жақты сараланып, филолог мамандар үшін түсінікті түрде баяндалады.

Кітап жоғары оқу орындарының оқытушыларына, аспиранттар мен ізденушілерге, магистранттар мен студенттерге және жалпы көпшілік қауымға арналған.

Лингвист-математик Асқар Жубанов вносит большой вклад в развитие прикладной лингвистики казахского языкознания. В данной книге собраны основные труды автора, где разрабатываются методология статоллингвистической, математической и компьютерной лингвистики, исследуются различные аспекты данной отрасли языкознания, его современное состояние и перспективы развития.

Книга излагается в доступной для специалистов-филологов форме и адресована преподавателям высших учебных заведений, магистрантам и студентам, а также для всех интересующихся читателей.

Ж  $\frac{44602020400}{00(05)-07}$

ББК 81.2-5

ISBN 9965-17-460-1

© Жұбанов А., 2008  
© “Арыс”, 2008

*Осы еңбегімді қазақ тіл білімінің  
классигі, ойшыл-филолог ғалым,  
1937 жылғы нәубеттің құрбаны —  
Әкем Құдайберген Қуанұлы ЖҰБАНОВТЫҢ  
ұлылығына бас ие отырып,  
әруағына бағыштадым...*

*Преклоняясь перед гениальностью  
Отца Кудайбергена Куановича ЖУБАНОВА —  
филолога-мыслителя, классика казахского  
языкознания, ставшего жертвой  
репрессий 1937 года, посвящаю  
свой скромный труд...*

**Автор**

## ҚАЗАҚ ТІЛТАНЫМЫНЫҢ ЖАҢА БІР САЛАСЫНЫҢ КӨШБАСТАРЫ

Асқар Жұбанов 1937 жылдың 13 қарашасында Алматы қаласында қазақ тіл білімінің теориялық негізін салушы атақты ғалым, тұңғыш кәсіби лингвист, түркітанушы, ағартушы, профессор Құдайберген Жұбановтың отбасында дүниеге келді. Бұл бүгінде баришамызға белгілі саяси қуғын-сүргіннің нағыз өршіп тұрған кезі болатын. Бір аптадан соң репрессия өрті ғалымды да жалмады, “жапон шпионы” деген жаламен түрмеге қамалып, артына миллиондаған сталиндік режимнің құрбандарының жолын құшты...

Шыр етіп жарық дүниенің есігін ашқаннан тағдырдың тепкісін көрген кенже бала Асқар өзінің екі аға, үш апасы, “халық жауының әйелі” атанған анасымен бірге Қазақстанның бір қиырынан екінші шетіне қоныс аударып, көшіп жүруге мәжбүр болды. Ата жұрт – Ақтөбе өңірі, Алматы, солтүстіктегі Петропавл, қайтадан Алматы деп, қашан ардақты әке есімі “ақталғанша” бір жерде тұрақтатпады. Алдымен Құдекеңнің – Құдайберген Жұбановтың әкелік мейірімі мен тәрбиесін, Ахаңның – академик Ахмет Жұбановтың қамқорлығы мен көмегін, жақсылығын көріп өскен, қала берді, қанға біткен қайсарлықтың арқасында әкеден шиеттей жетім қалған алты сарыауыз балапанның жерде қалғаны жоқ: отыз жылдай Алматыдағы сурет училищесіне басшылық еткен Ақырап, профессор Есет,

## ОСНОВАТЕЛЬ НОВОГО НАПРАВЛЕНИЯ КАЗАХСКОГО ЯЗЫКОЗНАНИЯ

Асқар Жұбанов родился в семье известного ученого, основателя теории казахского языкознания, первого профессионального лингвиста, тюрколога, просветителя, профессора Кудайбергена Куановича Жубанова 13 ноября 1937 г. в городе Алматы. Это было время, как известно нам из истории, свирепствования политической репрессии. Через неделю волна репрессии подхватила и ученого, по обвинению “японский шпион”, его посадили в тюрьму, после его наступила та же участь, что и других миллионных жертв сталинского режима.

Не успев появиться на свет, младший сын Асқар вместе с двумя старшими братьями, тремя сестрами и матерью, за которой прикрепили обвинение “жена врага народа”, были вынуждены скитаться с одного угла Казахстана на другой. До “оправдания” доброго имени отца они вынуждены были не задерживаться ни на одном месте: родной край – Актюбинск, Алматы, северный Петропавловск, затем снова Алматы. Успевшие повидать отцовское воспитание и нежность Құдекең – Кудайбергена Жубанова, испытать заботу и опеку Ахаң – Ахмета Жубанова, но рано потерявшие отца дети, несмотря на все тяготы жизни и трудности, выпавшую на их долю, стали достойными людьми: имя художника Ақырап, профессора Есет, автора “Букваря – Алиппе”, педагога Муслимы,



“Әліппенің” авторы, ұстаз Мұслима, экономист Қызғалдақ, Қырмызылар есімі еліміздің тарихымен бірге өріліп жатыр. Солардың бірі – бүгінде өке ісін жалғастырып, қазақ тіл білімінің тың саласының түрөнін тартып, тізгінін ұстап отырған профессор Асқар Жұбанов.

Мектеп қабырғасында нақты пәндерді ұнатқан жас Асқар оқу бітірген соң Абай атындағы Қазақтың педагогикалық институтының физика-математика факультетіне оқуға түсіп, 1960 жылы оны математика мамандығы бойынша үздік бітіріп шықты. Одан соңғы аралықтарда әскерлік борышын өтеп келіп, табан аудармай Қазақ мемлекеттік ауылшаруашылық институтының жоғары математика кафедрасында қызмет атқарады. Күндердің күнінде, отызға толған шағында көп ойланып, көп толғанып жаңа жолға, ғылымның соны соқпағына бет бұрады.

Бұл жылдар советтік тарихымызда “жылымық жылдар” атанып, ел еркін тыныстан, әдебиет пен өнер, ғылым мен техника жаңа сапада, тың серпінмен дами бастаған кезі еді. Дәстүрлі, классикалық ғылымдардың өзі жаңа өріс іздеп, пәнаралық, салааралық бұлақ көздері ашыла бастады. Сондай жаңа саланың бірі, өкінішке орай, тап бүгіннің өзінде қазақ топырағында өркенін кең жая алмай, көпшілік тосырқана қарап келе жатқан математикалық лингвистика еді. Бірде статоллингвистика, бірде компьютерлік лингвистика, кейде машиналық аударма сияқты аттарға ие болып, “қолданбалы лингвистика” атты ортақ шапан киген пәнаралық ғылымға бәріміз шекемізден қарап жүргенде бұл жасанды интеллектіңіз жекебас компьютері түрінде қағаз-қаламды айырбастап, бәріміздің төрімізден бір-ақ шыққанын да байқамай қалдық.

---

экономиста Кызғалдақ и Кырмызы, связано с историей нашего народа. Один из них – Аскар Жубанов продолжает дело отца, он занимается новым направлением в казахском языкознании.

Полюбив еще со школьной скамьи точные предметы, Аскар после окончания школы поступил на физико-математический факультет Казахского педагогического института имени Абая и в 1960 году успешно завершил его по специальности математика. Затем, отслужив в рядах армии, работал на кафедре высшей математики Казахского государственного сельскохозяйственного института. К тридцати годам он вышел на новый путь науки.

Этот советский период нашей истории считался “периодом оттепели”, когда народ мог свободно дышать, литература и искусство, наука и техника стали развиваться в разных направлениях. Традиционные классические науки стали искать новые междисциплинарные, межотраслевые подходы. Одним из этих направлений, к сожалению, не нашедших в данный момент широкого применения, является математическая лингвистика. В начале междисциплинарная наука получившая название статоллингвистика, машинный перевод, затем компьютерная лингвистика, объединенная под общим названием “прикладная лингвистика” обрела почетное место, связано было это заменой пера и бумаги на искусственный интеллект – компьютер.

В эти годы в Москве и Ленинграде, а также в Минске стали создаваться исследовательские центры, связывающие лингвистику с математикой.

Осы жылдар Мәскеу мен Ленинградта, сондай-ақ Минскіде лингвистика мен математиканы ұштастыра зерттейтін орталықтар құрыла бастады. Оның бастауында профессор Р.Г.Пиотровский тұрған еді. Қашанда ғылымда жаңа нәрсе тез таралады. Қазақстанда шымкенттік математик (кейіннен тілші ғалымға айналды) Қалдыбай Бектаев осы іске бел шеше кірісті. Болашақты болжай білген Қазақ ССР Ғылым академиясының басшылары, оның ішінде Тіл білімі институтының директоры Исмет аға Кеңесбаев уақтылы осы бастамаға үн қоса білді, ең алғашқы одақтас республикалардың бірі болып Академиялық институт жанынан арнайы ғылыми топ құрылды. Міне, осы жаңа істің басы-қасында алғашқылардың бірі болып “пионер” Асқар Жұбанов та жүрді.

Әрине, математик Асқардың тілшілік (математикалық-тілшілік, болмаса тілшілік-математикалық) жолды таңдауы оңай болмады. Алдымен әке жолын қуып туабітті тілші немесе әдебиетші бола білген, сөзі ұзын, қаламы жүйрік Есеттей ағасының үгіт-насихаты бір себепші болса, екіншіден, көзі көрмесе де құлағы қанық ғалым әкесін арбаған тіл ғылымының тұңғығына өзі де ең болмағанда бір рет сүңгіп шығуға аңсары ауғаны кәдік. Сөйтіп, 1967–1970 жылдары қазақ тілін кешенді зерттейтін әлемдегі жалғыз қара шаңырақ — Қазақ ССР Ғылым академиясының Тіл білімі институтының күндізгі бөлімінің аспиранты болады және осы үш жылын одақтық деңгейдегі математикалық лингвистиканың үлкен орталығы — Минск қаласындағы Шет тілдер институтында өткізді. Осы жерде өзі секілді аспирант, бірақ кейін ғылым жолында ұстазындай бола білген А.В.Зубовпен табысып, күні бүгінге дейін араласып, ақылдасып тұрады. Ол жылдарда

---

Основателем был профессор Р.Г. Пиотровский. В науке новое всегда быстро находит распространение. В Казахстане известный математик-вероятностник (после стал языковедом) Калдыбай Бектаев вплотную стал заниматься этим направлением. Сумевшие предвидеть перспективу данного направления руководство Академии наук КазССР, среди них директор Института языкознания Смет аға Кенесбаев вовремя поддержало это начинание, среди союзных республик первыми организовали специальную научную группу при Академическом институте. Среди всех принимал в этом участие и “пионер” Аскар Жубанов.

Конечно, для математика Аскара выбрать путь языковеда было не легко. Но несмотря на это, во-первых, было влияние брата Есета, который сумел стать хорошим языковедом и литератором, как и отец, во-вторых, он сам также мечтал окунуться в науку, которой занимался его отец. Таким образом, в 1967–1970-е годы он становится аспирантом дневного отделения единственного в мире очага, занимавшегося исследованием казахского языка — Института языкознания Академии наук КазССР и эти три года он провел в Институте иностранных языков в городе Минске — большом математико-лингвистическом центре на уровне союза. Здесь он знакомится с таким же аспирантом, после ставшим учителем, А.В. Зубовым и до сих пор поддерживает с ним связь. В те годы основным средством статоллингвистики был ЭВМ (электронно-вычислительная машина), который, естественно, “не говорил” на казахском. Аскар Жубанов, как

статоллингвистердің негізгі құралы — ЭЕМ (электронды есептеуіш машина), оның өзі қазақша “сөйлеп” көрген емес. Оған арнайы бағдарлама түзіп, түрлі сөз табына жататын сөздерді, сөз қолданыстарды машина көмегімен анықтап, неше түрлі жиілік, кері алфавиттік сөздіктер түзу сияқты тың істердің басында тұрған, тура мағынасындағы “пионер” ғалым Асқар Жұбанов болатын. Осылай тыңға түрен сала жүріп ол 1973 жылы кандидаттық жұмысын табысты қорғады және осы жолдағы ұзақ жылғы ізденістерін “Квантативная структура казахского текста” деген атпен (“Ғылым”, 1987) жарыққа шығарды. Осылайша қазақ ғалымы қазақ тілі мәтіндерін ЭЕМ-нің жасанды миында “қайналып” алып, алғаш рет ана тіліміздегі тілдік бірліктердің (сөзформа, негіз сөз, өріп, өріп тіркесі т.б.) автоматтанған жиілік сөздіктерін құрастырудың алгоритмі мен бағдарламасын жасап шығып, оны іске қосты. Сонымен бірге қазақ мәтініне сүйене отырып жалпы түркі тілдері мәтіндерінің ықтималдық-статистикалық моделін жасап шығудың алғашқы нәтижелеріне қол жеткізілді. Математикалық статистика мамандары ғана түсіне алатын Нормаль, Пуассон, Шарлье үлестірілімдік заңдарына сүйене отырып, қазақ мәтініндегі зат есім мен етістіктердің алғашқы заңға бағынатынын, көпшілік жағдайда үстеу сөздердің Пуассон заңы мен Шарлье заңының В түріне байлаулы екеніне алғаш рет көз жеткізілді. Осылайша қазақ мәтініндегі мыңдаған сөзқолданыстардың бір қазанда кездейсоқ бас қосып, қайнаған материалдар емес, сөз таптарына ажыратылу, үлес салмағы тұрғысынан белгілі бір ықтималдық-статистикалық заңдылыққа бағынатыны дәлелденді.

истинный “пионер” этого дела, стоял у истоков создания специальной программы, которая дает возможность с помощью ЭВМ создавать различные частотные словари, как обратно-алфавитные частотные словари, которые дают возможность определить слова, относящиеся к разным частям речи. Плодотворно трудясь, он в 1973 году успешно защитил кандидатскую диссертацию и выпустил на свет свой труд “Квантативная структура казахского текста” (“Наука”, 1987). Таким образом, казахский ученый ввел в память ЭВМ тексты, относящиеся к различным стилям казахского языка и впервые составил на ЭВМ алфавитно-частотные словари различных языковых единиц (словоформа, основа, буква, сочетание букв и др.). Вместе с тем, основываясь на полученные данные по текстам казахского языка, смог добиться определенных результатов по созданию вероятностно-статистических моделей тюркских текстов. Опираясь на Нормальный закон распределения, на законы распределения Пуассона и Шарлье, понятные только математикам, впервые определил, что в казахском тексте имя существительное и глагол подчиняются Нормальному закону, а наречие в основном подчиняется законам Пуассона и Шарлье типа-В. Тем самым было доказано, что использование слов в казахском тексте это не просто набор слов, а совокупность текстовых единиц, распределенных в тексте, подчиняясь определенным вероятностно-статистическим законам.

В своей кандидатской диссертации ученый определяет статистическую структуру эпопеи “Путь Абая” М. Ауэзова. После, трудясь плодотворно в этом

Ғалым осы кандидаттық жұмысының өзінде М.Әуезовтің “Абай жолы” эпопеясының статистикалық құрылымын анықтауға ұмтылды. Кейін осы салада өзімен бірге өнімді еңбек еткен ғалымдармен бірлесе отырып “М.Әуезовтің “Абай жолы” романының жиілік сөздігін” түзіп шықты (“Ғылым”, 1979, 336 бет). Осы еңбектің нәтижесінде классик жазушымыздың эпопеяда 465691 сөзқолданысты пайдаланғанын, ондағы сөз саны — 16983, ал сөзформасы — 61424 екені анықталды. Сонымен бірге белгілі түркітанушы ғалым Ә.Құрышжановпен бірлесе отырып жарыққа шығарған “Куманиша-қазақша жиілік сөздік” те (“Ғылым”, 1978, 216 бет) түсіне білген жанға сандық көрсеткіш арқылы тілдік сапаны жарқырата ашып көрсететін күрделі еңбек-тұғын. Өйткені әлі күнге ана тілімізге толықтай аударылмай өрі асыл нұсқасы жарық көре алмай келе жатқан “Кодекс Куманикс” (XIII ғ.) — қазақ тіліне тікелей қатысты көне қыпшақ жәдігері.

Ғалым 1997 жылы тағы бір жыл бойы Белоруссия Ғылым академиясының Тіл білімі институтында докторантура есебіндегі ғылыми іссапарда болады. Профессор А.В.Зубовтың кеңесшілігімен ұзақ жылдарғы зерттеуінің қорытындысы есепті “Основные принципы формализации содержания казахского текста” деген тақырыпта монографиясын жарыққа шығарып (Алматы, 2002, 250 бет), докторлық диссертациясын сәтті қорғап шықты. Тіл білімінде соңғы жылдары ғана алдыңғы қатарға шыққан жаңа зерттеу бағыты — мәтін лингвистикасы тұрғысынан жазылған, мәтіннің мазмұндық құрылымының түзілуін абзацтық деңгейде қарастырған күрделі еңбектің негізгі мақсаты — қазақ мәтінінің семантикалық көрінісін бейнелеу және оның туындау мүмкіндігінің формалды принциптерін анықтап, көрсе-

---

направлении, вместе с другими учеными, создал “Частотный словарь романа М. Ауэзова “Путь Абая” (“Наука”, 1979, 336 с.). В результате этого труда удалось определить, что писатель-классик использовал в эпопее 465691 словоупотреблений, из них 16983 — слова, а словоформы — 61424. Кроме того, созданный совместно с известным тюркологом, ученым А. Курышжановым “Частотный куманско-казахский словарь” (“Наука”, 1978, 216 с.) является сложным трудом, который может раскрыть качественное использование языковых единиц с помощью статистических данных. Этот труд является важным и ценным, так как до сих пор полностью не переведенный на казахский язык “Кодекс Куманикс” (XIII в.) — кыпчакский памятник, имеет прямое отношение к казахскому языку.

Ученый в 1997 году был в командировке по докторантуре в Институте языкознания Академии наук Белоруссии. В результате полученных консультаций от профессора А.В.Зубова выпустил в свет свой труд “Основные принципы формализации содержания казахского текста” (Алматы, 2002, 250 с.), как итог многолетней работы, успешно защитил докторскую диссертацию. Основная цель ценного труда, который рассматривает содержательную структуру текста на уровне абзаца, написанный по новому направлению языкознания — лингвистике текста — описание семантической картины казахского текста и определить, выделить формальные принципы их порождения. Таким образом, мы можем смело сказать, что это были попытки применить современные информационные технологии для расширения рамки использования государственного языка. Предлагаемая ученым

ту. Бір сөзбен айтқанда, бұл мемлекеттік мәртебеге ие болып отырған ана тіліміздің қазіргі заманғы ақпараттық технологиялардың жаңалықтарына жедел бой үйретіп, соның нәтижесінде қолданыс аясының кеңеюіне ықпалдастық жасау. Атап айтқанда, зерттеуші ұсынып отырған ана тіліміздің формалданған моделі машиналық аударма тәжірибесінде, тіл оқытудың компьютерлік бағдарламасын жасауда, қазақ лексикографиясын жаңа деңгейге көтеруге, яғни оны автоматтандыруға және т.б. тек теориялық тұрғыдан ғана емес, осындай қыруар қолданбалы, практикалық мәселелерді шешуде де пайдасы шаш етекпен болмақ.

Міне, осылайша осыдан 40 жыл бұрын тілшілер ауылына қорқасоқтап, қымсына аяқ басқан талапты жас бүгінде қазақ тіл білімінің жаңа да болашағы зор “қолданбалы лингвистика” саласының көшбастаушы “ақсақалы” дәрежесіне абыроймен жетіп отыр. “Шөп те шыққан жеріне шығады”, 40 жылдан бері әке жолын жалғап, табан аудармай А.Байтұрсынұлы атындағы Тіл білімі институтында қызмет атқарып келеді. Аспирант, кіші ғылыми қызметкер, 1976 жылдан аға ғылыми қызметкер, 1986 жылдан жетекші ғылыми қызметкер, ұзақ жыл “Қолданбалы лингвистика” бөлімін басқарды (1990–97). Соңғы онжылдықтағы түрлі реформа, өзгерістерге байланысты Институтта өз саласы бойынша бас ғылыми маман ретінде қызмет атқарып келеді.

А. Жұбановтың тікелей атсалысуымен Институтта соңғы жылдары атқарылып келе жатқан үлкен жоба — “Қазақ тілінің компьютерлік қоры”. Мұнда ана тіліміздің бар байлығы, ақпараттанған білім қоры жинақталған. Соңғы кезде “Тіл—Қазына” атауына ие болған картотекалық қордың компьютерлік базасын асыл

---

формальная модель родного языка может быть использована в переводческой деятельности, в создании компьютерной программы по обучению языку, для автоматизации задач, относящихся к лексикографической отрасли казахского языкознания, а также в решении не только теоретических, но и ряда практических проблем.

Вступивший 40 лет назад на путь языковеда молодой ученый на сегодняшний день справедливо занял почетное место “аксакала” среди основоположников “прикладной лингвистики”. В течение 40 лет по стопам отца он преданно служит в Институте языкознания имени А.Байтұрсынова: аспирант, младший научный сотрудник, с 1976 года старший научный сотрудник, с 1986 года ведущий научный сотрудник, долгие годы руководил отделом “Прикладная лингвистика” (1990–97). В последние десятилетия в связи с различными реформами в Институте работает как главный научный сотрудник.

По инициативе А. Жубанова в Институте в последние годы работают над большим проектом — “Компьютерная база казахского языка”. Здесь собрано все богатство, информационный фонд языка. Получившее название “Тіл — Қазына” — картотечный фонд компьютерной базы можно назвать “Банком данных о казахском языке”, где хранятся все богатство родного языка, культурное наследие. В будущем новая волна молодых ученых, которые в совершенстве будут владеть компьютером, смогут использовать это богатство и внесут большой вклад в развитие казахской лексикографии.

сөз, мәдени мұрамыз сақталатын “Тілбанкі” десек болғандай... Таяу болашақта компьютердің құлағында ойнайтын жаңа толқын ғалымдар ол байлықты сарқа пайдаланары, әсіресе кенжелеп қалған лексикографиямыздың өркендеуіне қосар үлесі қомақты болары сөзсіз.

Профессор А. Жұбановты соңғы жылдары Алматыдағы жетекші білім ошақтары — әл-Фараби атындағы Қазақ ұлттық университеті мен Абай атындағы Алматы ұлттық педагогикалық университеттері аттай қалап шақырып, қолданбалы лингвистиканың түрлі саласынан арнайы пәндер ашып, студенттеріне дәріс тыңдатып жүр. “Ұстаздық еткен жалықпас” дегендей, бұл орайда да ғалым ісі өнікті. Сол жерден өзіне бірнеше ізбасар, шәкірт тапты, алды кандидаттық диссертацияларын қорғап, кейбірі ғылым жолына, тың салаға атбасын бұрып жатыр. Бұған дейін басты еңбектерін орыс тілінде жазып келе жатқан А. Жұбанов осы университет қабырғасында оқыған дәрістері негізінде қолданбалы лингвистиканы ана тілімізде сөйлетіп, оқулықтардың тұтас сериясын жасап шығаруды мақсат тұтқандай. Оған мына қолымызға ұстап отырған оқу құралдары куә: “Компьютерлік лингвистикаға кіріспе” (Алматы, 2007), “Қолданбалы лингвистика: қазақ тілінің статистикасы” (2-бас., Алматы, 2007), “Қолданбалы лингвистика: формалды модельдер” (2-бас., Алматы, 2007).

Адам өзі өрекіні көрсетпесе де, даурығып дәлелдемесе де, ең мықты ғалым оның бойында жатқан небір жақсы қасиеттері барлардан шығатын сияқты. Жұбановтар әулетінің қай өкіліне қарасақ та, үлгі тұтарлық кісілік, ұстамдылық, талаптылық пен таланттылық, еңбекқорлық қасиеттерді көреміз. Өздерін жоғары орынға апарған да, біздерді сүйсіндірген де осы қасиеттер болар.

---

*В последние годы профессора А. Жубанова приглашают читать лекции студентам по прикладной лингвистике ведущие вузы Алматы, такие как Казахский национальный университет имени аль-Фараби и Алматинский национальный педагогический университет имени Абая. Благодаря своему трудолюбию и в этом деле он имеет большие успехи. Там он нашел учеников, некоторые из них на стадии защиты кандидатской диссертации, а другие только начинают свой путь в науке. До этого времени писавший свои труды на русском языке А. Жубанов на основе прочитанных лекций поставил цель выпустить учебные пособия по прикладной лингвистике на родном языке. Этому могут послужить доказательством такие учебные пособия, как: “Компьютерлік лингвистикаға кіріспе” (Алматы, 2007) — Введение в компьютерную лингвистику, “Қолданбалы лингвистика: қазақ тілінің статистикасы” (2-изд., Алматы, 2007) — Прикладная лингвистика: статистика казахского языка, “Қолданбалы лингвистика: формалды модельдер” (2-изд., Алматы, 2007) — Прикладная лингвистика: формальные модели.*

Для человека не обязательно доказывать, что он выдающийся ученый, если в нем это заложено. Каждому представителю Жубановской династии присущи человечность, сдержанность, требовательность и талантливость, трудолюбие. Эти качества дали возможность достигнуть высшего уровня, и они же заставляют нас уважать их.



Өмірде кездесіп жататын өлім-қаза, ауру-сырқау сияқты ауыр жүктерді көтере білу де адамға үлкен сын. Осының бәрі нар көтерер жүк болса, сол жүкті де Асқар көтеріп келеді. Заман мен солақай саясат сынынан өте білген ғалым-азамат тағдыр сынына да арқасын тосып отыр. “Өмірдің жақсы да, жаман да жақтарын басымнан кешірдім, бірақ оның қиын тұстарының бәрі де әкемнің нақақтан “халық жауы” аталуымен салыстырғанда түк те емес сияқты” — дейді ғалым. “Асылдың сынығы, тұлпардың тұяғы”, өзі де сом асыл — мықты ғалым, атаулы азамат, үлгілі ұстаз, филология ғылымдарының докторы, профессор Асқар Құдайберген баласы, халқымыз мақтанышпен қастерлеп атайтын “Жұбановтар әулеті” деген қымбат атты жоғары ұстап, кейінгі ұрпақтарға жеткізе бермек, әлі де том-том еңбек жазып, топ-топ шәкірт дайындай бермек!

**Рәбиға СЫЗДЫҚ,**  
академик;  
**Ғарифолла ӘНЕС,**  
филология ғылымдарының  
кандидаты

---

*В жизни бывает много трудностей, перенести их — это огромное испытание для человека. Немало испытаний выпало и на долю Асеке, которые он сумел мужественно победить. “В жизни видел и хорошее, и плохое, но все они не сравнятся с тем, что отца обвинили ложно “врагом народа”, — говорит ученый.*

*Сын великого ученого Аскар Кудайбергенович сам стал видным лингвистом, образцовым педагогом, высоко держащим имя династии Жубановых, которого чтит наш народ. Ждем от него дальнейшей плодотворной работы в научной и педагогической деятельности.*

**Рәбиға СЫЗДЫҚ,**  
академик;  
**Ғарифулла АНЕС,**  
кандидат филологических  
наук

# ҚОЛДАНБАЛЫ ЛИНГВИСТИКА: ҚАЗАҚ ТІЛІНІҢ СТАТИСТИКАСЫ



*Кітапта қолданбалы лингвистика саласы бойынша қазақ тілін зерттеу нәтижелеріне қысқаша ғылыми шолу жасалып, жиілік сөздіктер мәселесі және олардың тілді зерттеу ісінде пайдаланылуы, автоматты түрдегі жасалу жолдары көрсетіледі. Тіл зерттеудегі статистикалық әдістің бастама түстары сипатталып, негізгі сөз таптарының қазақ мәтініндегі үлестірілуінің ықтималды-статистикалық үлгісін (модельін) құрудың проблемалық мәселелері қарастырылды. Тілдің лексика-морфологиялық құрылымына статистикалық әдісті қолданудың алғышарттары берілген.*

- АЛҒЫ СӨЗ
- КІРІСПЕ
- *Бірінші тарау*  
ЖИЛІК СӨЗДІКТЕР
- *Екінші тарау*  
СТАТИСТИКАЛЫҚ ЛИНГВИСТИКА
- *Үшінші тарау*  
ҚАЗАҚ МӘТІНІН ЫҚТИМАЛДЫ-СТАТИСТИКАЛЫҚ МОДЕЛЬДЕУ
- *Төртінші тарау*  
ТІЛДІҢ ЛЕКСИКА-МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫНА  
СТАТИСТИКАЛЫҚ ӘДІСТІ ҚОЛДАНУДЫҢ АЛҒЫШАРТТАРЫ
- ҚОРЫТЫНДЫ
- ҚОСЫМША
- ӘДЕБИЕТ

## АЛҒЫ СӨЗ

Оқырмандар назарына ұсынылып отырған бұл еңбек\* қолданбалы лингвистиканың бір тармағы болып саналатын статистикалық лингвистиканың кейбір өзекті мәселелерін сөз етеді. Автор аса көрнекті ғалым Бодуэн де Куртенэнің: “Тіл білімінде сандық, математикалық ойлауды жиі қолданып, оны мейлінше нақты ғылымдарға жақындату керек” деген ұстанымын басшылыққа алады [22]. Сондай-ақ Р.М. Фрумкинаның “Статистические методы при заведомо неполной информации об объектах исследования предоставляют возможность делать выводы об этих объектах с заданной точностью и надежностью” [91] — деген пікірімен толық келісе отырып, тілге қатысты көптеген тұжырымға статистикалық ғылыми баға беріп, олардың дұрыс-бұрыстығын тексеруге болатындығына арнайы тоқталады.

Ықтималды-статистикалық заңдылықтар табиғат пен қоғамда кездесетін әр алуан құбылыстардың өзара қарым-қатынастарында көрініс табатыны белгілі. Осындай көрініс “жиілікпен” өлшенеді десек, артық айтқандық болмас. Ал тілдік құбылыстың жиілігі — әмбебап тілдік категория. Бұл жерде тіл зерттеу барысында тек математикалық, дәлірек айтсақ, статистикалық әдіс-тәсіл мен математикалық өрнекті (формуланы) қолдану туралы ғана сөз болып отырған жоқ, сонымен бірге сандық сипат белгілі бір тілдің ішкі табиғатына объективті түрде тән деп түсінген жөн. Шынында да, бүгінгі таңда тіл құрылымындағы сапалық және сандық сипаттар іштей тығыз байланыста болатыны ешқандай күмән туғызбайды. Мәселен, кейбір беделді, ірі ғалымдар өз тұжырымының дұрыстығына толыққанды дәлелдер келтіруі мүмкін. Осыдан тіл білімінде не басқа ғылым салаларында бірін-бірі терістейтін теориялар мен болжамдар (гипотезалар), тіпті ғылыми мектептер пайда болады. Міне, сондықтан тілді статистикалық тәсілдермен зерттеу тілші-мамандардың субъективті тұжырымдарын не объективті шындыққа айналдыруға көмектеседі, немесе оның теріс екендігін дәлелдеп береді.

Кітап тілдің сөздік қорын зерттеуді мақсат еткен филолог-студент, филолог-мұғалім, филолог-зерттеуші оқырмандарды ең қарапайым түрдегі статистикалық әдістермен және оның қарапайым есеп-қисап жүргізу жолдарымен таныстыруды көздейді.

Сөздік қор — аса күрделі тілдік нысан, бірақ көптеген теориялық және практикалық мәселелер үшін оның кейбір қасиеттерін, яғни ең негізгі деген сандық сипаттарын білу жеткілікті. Мәселен, кейбір жағдайда белгілі көлемдегі мәтін мен оның сөздігі арасындағы жоғары ықтималдыққа ие қатынастарды білсек те жеткілікті екен. Мәтін көлемінің өзгеруіне қарай зерттеу нысанына алынған тілдік бірлік (сөз не сөз тіркес) жиіліктерінің мәтін мен сөздік бойында таралу (үлестірілу) сипаты қандай, алынған сандық нәтижелерді ана тілін, не шет тілін үйретуде қалай қолдануға болады — деген сауалдарға жауап алуға мүмкіндік береді. Сонымен

\* Бұл оқу құралына автордың орыс тілінде жарық көрген “Квантитативная структура казахского текста” атты монографиясы негіз болды [2]. Бірақ аталған құрал осы еңбектің тікелей аудармасы емес. Өйткені оқу құралы мен ғылыми монографияның құрылымдары мен мақсаты әртүрлі және 1987—2004 жылдар аралығындағы ғылыми жаңалықтар ескеріліп, оқу құралына тиісті толықтырулар енгізілді.

қатар тілдік ақпаратты тарату, ғылыми ақпараттарды автоматты түрде өңдеп, ұзақ мерзімге сақтау, оларды қажеттігіне қарай іздеп, тауып алу мүмкіндігін анықтау да сөздік қордың статистикасын зерттеу арқылы іске асады.

Қолыңыздағы оқу құралына қолданбалы лингвистика саласының “статистикалық тіл білімі” негізгі нысан болуының өзіндік себептері де бар.

Кезінде Бүкілодақ бойынша “Тіл статистикасы” 1957 жылдан зерттеле басталса, филология ғылымдарының докторы, математик Қ.Б.Бектаев жетекшілік еткен қазақстандық ғылыми шығармашылық топ зерттеу жұмыстарын 1970 жылдан бастап жүргізе бастағаны ғылыми көпшілікке мәлім. ҚазССР ҒА Тіл білімі институтындағы “Статоллингвистикалық зерттеу және автоматтандыру” деп аталатын ғылыми шығармашылық топ қазақ әдебиетіндегі белгілі ақын-жазушылар шығармалары мен басқа да стильдерге қатысты мәтіндер тілін статистикалық әдіспен зерттеді. Содан бері осы сала бойынша едәуір ғылыми еңбек жарық көрді. Қазақстандық статоллингвист ғалымдардың жұмысына қысқаша шолу жасаумен қатар біз өз зерттеу тәжірибемізді де оқырмандарға таныстыруды мақсат етеміз.

Тілдегі заңдылық жүйеге ғана тән болғандықтан, статистикалық заңдылықтарды ашуға байланысты қолданылатын әдістер кезінде тілдің өзінше бір бөлек жүйе құратындығы туралы сөз ете бермейміз. Басқаша айтсақ, тіл – біріне-бірі қатыссыз құбылыстар жиынтығы емес, керісінше, заңды түрде ұйымдасқан, көпөлшемді және тәуелсіз сипаттағы ақиқат.

Тілдік бірліктердің (дыбыс, морфема, сөз тіркестері, сөйлем) мәтін ішінде (жазба тіл ағымында) кездесуі кездейсоқ оқиға деп есептесек, онда белгілі бір жүйеден туындайтын тілдің өзі де (яғни мәтін) кездейсоқ процеске қатысты болады. Осының салдарынан зерттеу нысанына алынатын нақты бір мәтін белгілі дәрежеде ақиқат сөйлеу тіліне жақын келетін модельденуші мәтінге айналады. Мұндай жағдайда мәтін қасиеттері тұрақты және элементтері бір-біріне тәуелсіз деген қағидатты ұстану қажет болады.

Лингвистикалық статистикада тілдік заңдылықтарды анықтау барысы тәжірибелік байқау мен сол байқаудың негізінде туындайтын нәтижелерді математикалық жолмен өңдеу арқылы іске асады. Сондықтан лингвостатистикалық зерттеулер жаратылыстану ғылымдарындағыдай табиғи және ғылыми тұрғыда жүргізіледі.

Статистикалық деректерді өңдеуде компьютерлік технологияны пайдалану, тілдік мәселелердің шешімін тез арада табу, математикалық статистика әдістерін басқа ғылымдар тәсілдерімен (ақпараттық теория, математикалық логика және т.б.) ұштастыру казiргi кездегi отандық және шетелдiк тiл бiлiмiнiң даму бағытына сай келедi.

Түркі мәтіндерін статистикалық тәсілдермен зерттеу, ғалымдардың пайымдауынша, бұл тілдердің ішкі табиғатына, яғни олардың жалғамалық (агглютинативтік) құрылымына негізделеді.

Тіл зерттеу тәжірибесінде сандық зерттеулердің қажеттігін математикалық лингвистика мамандарымен қатар дәстүрлі тіл білімінің аса көрнекті өкілдері де мойындап келеді. Олардың пікірінше, жазба әдебиет пен сөйлеу тіліндегі түрлі сөздердің әр ыңғайдағы қолданылу жиілігін анықтау тәсілі тілдің құрылымды-грамматикалық, кейде стильді-семантикалық айырымдарын ажыратумен қатар стильдерге тән қасиеттерді де анықтауға көмектеседі.

Қорыпа айтқанда, барлық грамматикалық категорияларды сандық тәсілмен талдау олардың әдеби тілдегі функционалдық салмағын нақты көрсетуге мүмкіндік береді.

Осы оқу құралының қолжазбасымен танысып, құнды-құнды пікір айтқан әл-Фараби атындағы Қазақ ұлттық университеті жалпы тіл білімі кафедрасының меңгерушісі профессор Ә.Д.Сүлейменоваға және осы кафедраның профессор-оқытушы құрамына автор өз ризашылығын білдіреді.

## К І Р І С П Е

Тілдің қолданбалы аясы бұрыннан да өзінің кең және жан-жақтылығымен ерекшеленетін. Оның ескіден келе жатқан саласы – жазу (графика), ана тілі мен шет тілін оқыту әдістемесі және тілдің лексикографиялық жүйесі. Осылардан кейін барып – аударма, шифртану (дешифровка), орфография, транслитерация және терминологияны өңдеу. Қолданбалы лингвистиканың дәстүрлі бағыттарының бірі – мемлекеттің тіл саясатына қатынасу:

- 1) әліпбиді, орфографияны өңдеу, сауатсыздықты жою;
- 2) мемлекеттік тілді тағайындау;
- 3) мемлекеттік тілден басқа тілдердің орнын анықтау;
- 4) ұлттық терминологияны өңдеу, бірізге келтіру және қалыптастыру;
- 5) қала, көше, алаңдарға ат қою мен оның бұрыннан келе жатқан атын өзгерту, яғни ономастика мәселелері.

Қолданбалы тіл білімінің осындай классикалық бағыт-бағдарын дамыту және мейлінше жетілдіру мәселелерімен бірге ХХ ғасырдың екінші жартысынан бастап оның бірқатар жаңа бағыттары орын ала бастады. Бұл бағыттар заманға сай қоғамдық, жаратылыстану және техникалық ғылымдардың өзара даму сипатынан туындайтын қолданбалы лингвистиканың тарихи дамуының логикалық жалғасының көрінісі еді.

Негізінен алғанда, адам баласының қызметінің өртүрлі аясындағы жұмыстарын жеңілдету бір ғана проблемалық мәселеге тіреледі. Ол – қоғам өміріндегі ақпарат жұмысын өңдеу. Бұндай ақпарат мәтін түріндегі жазба тілінде не үйреншікті сөйлеу тілінде болуы мүмкін.

Қоғамның практикалық қажеттігін өтеуде, әсіресе, ақпаратты сақтау мен оны тарату істерінің сырын ашуда зерттеуші оны ең алдымен тілдің ішкі табиғатындағы заңдылықтардан іздегені жөн.

Осымен байланысты өзекті мәселенің бірі – оператордың дауысымен басқарылатын станок пен құрал-аспаптарды өндіру. Ол үшін өртүрлі тілдердегі ақпараттарды автоматты түрде өңдеу мен оларды іздестіру, байланыс торабы жұмысын (телефон, радиобайланыс т.б.) жетілдіру, адам баласының сөйлеу мен есту қабілетінің бұзылуына қатысты ауруларды емдеуде дыбыстың фонетикалық деректеріне жүгіну және т.б.

Аталған проблемалық мәселелерді ЭЕМ-ды (компьютерді) кең түрде қолдана отырып, мәтінді (жазба, сөйлеу) автоматты түрде өңдеу қажеттілігі туындайды. Мәселен, әр типті ақпараттық жүйелерді тілдік деректермен қамтамасыз ету; машиналық аударма мәселесі; табиғи тілді түсінетін жүйе құру (жасанды интеллект жүйесіндегі тілдік мәселелер); сөйлеу кезіндегі дыбыстық сигналдарда бар ақпараттарды пайдалану негізінде арнайы жүйе құру т.б.

Соңғы жылдары өртүрлі ақпараттық жүйелерді тілдік деректермен қамтамасыз етумен байланысты терминтану мәселесін бірізділікке түсіріп, тұрақтандыру өзекті



мәселеге айналып отыр. Себебі, әр білім саласының ғылыми және техникалық терминдерді қолдану қажеттігі барынша сұраныс тудыруда. Әрине, мұндай мәселе лексикография саласынан тыс өз шешімін таппайтыны белгілі. Сондықтан қоғам өмірінде ЭЕМ-нің (компьютердің) кең қолдануымен тығыз байланысты автоматты лексикография саласы біртіндеп өз отауын құруда.

Практикалық мәселелерді шешуге байланысты табиғи тілде диалогты іске асыру үшін адам мен ЭЕМ арасында әсерлі қарым-қатынас орнату қажет. Мәселен, оларға жататындар: сұрақ-жауап жүйесін құру, роботтарды басқару жүйесін құру, аталған басқару жүйелерінде дұрыс шешім қабылдау үшін диалогтық процесті іске қосу. Мұндағы ең негізгі мәселе — автоматты құрылғының жазба не дыбыстық мәтінді түсіне білу (тану) проблемасы.

Әрбір тілдің қолданбалы мәселені шешуде өзіне тән ерекшелігі бар. Мұндай есептердің санын, түрін алдын ала білу мүмкін емес. Өмірде олар жыл сайын, ай сайын ауысып, біреуі келіп, екіншісі кетіп жатады. Қолданбалы лингвистикаға қатысты іргелі зерттеулер ғылыми-техникалық, ұйымдастыру-басқарудың ауызша-жазбаша құжаттарының, сөздік түзу мен және т.б. түрлерінің фонетикалық, грамматикалық, семантикалық және статистикалық құрылымын сипаттау мен модельдеу жақтарын қамтиды. Жекелеп айтсақ, оған теориялық және қолданбалы лингвистиканың шекарасында жатқан мәтіндік бірліктердің формальды моделін (үлгісін) құру проблемасын жатқызуға болады.

Теориялық лингвистика, негізінен, тілді қалыптық, жүйелік тұрғыда қарастыратындықтан, қолданбалы тіл білімі тілді әрекет үстінде, яғни оның қарым-қатынас (коммуникация) кезіндегі табиғатын түсінуге тырысады. Көптен бері қолданбалы тіл білімі ғылыми-техникалық пен іскерлік прозадан өзін алшақ ұстағаны аян. Дегенмен, ХХ ғ. 70-жылдарында ғалымдардың мынадай шешім жасауына тура келді. Көптеген қолданбалы мәселелер таза лингвистикалық емес, тіпті дәлірек айтсақ, ол проблеманы шешу адамның іс-әрекеті мен ойлау процесін, тілдің семантикасы мен оның формальды және семантикалық әдіс-амалын синтездеу (жинақтау) жолдарын модельдеуге тікелей қатысты. Осыдан барып қолданбалы тұрғыдағы зерттеулердің іргелі проблемасы — білімді модельдеу (моделирование знаний) мәселесі анықталды.

Соңғы кезде бұл проблема (білімді модельдеу) бірнеше ғылымдар тоғысына, дәлірек айтқанда — логика, лингвистика, математика, психология, кибернетика салаларының аясына қатысты болып отыр. Көптеген ғылымдардың осы мәселеге қызығушылығының негізгі себебі біреу-ақ, ол — автоматты жасанды интеллект жүйесін құру. Солай бола тұра, “білім” бізге “тіл” арқылы берілетіні анық. “Білім” сөйлеу мәтінінде де (монолог, диалог, әр жағдайдағы реплика) және сонымен бірге жазба мәтіндерде де (ескі жазба ескерткіші, көркем әдебиет, ғылыми-техникалық әдебиет) көрініс табады. Тіл арқылы “білімді” біз ұрпақтан-ұрпаққа жеткіземіз. Сондықтан “тіл” — “білімді” сақтау формасы және оны тарату құралы десек те болады. Біздіңше, ғылым мен техникада мәтіннен тыс жатқан “білім” жоқ деуге болады, ал ғылыми-техникалық мәтіннің семантикасын модельдеу дегеніміз — ол осы салалардың білім жүйесін модельдеу. Осылайша, біртіндеп келіп, білімді модельдеу проблемасы мәтін мазмұнын (мағынасын) модельдеумен ұштасып жатқанына көз жеткізуге болады. Бұл жерде ең негізгі шешімін табуы қажет ететін мәселе — мәтіннің семантикалық көрінісін (семантическое представление) құру.

Дегенмен, техникалық және іскерлік коммуникация аясындағы нақты қолданбалы мәселені шешу сол саладағы құжаттық мәтіндерінің грамматикалық, лексикалық, семантикалық құрылымының сипатталуына байланысты. Сонымен бірге, аталған салалардың терминологиялық сөздігінің болуына, мәтіннің статистикалық құрылымының зерттелуіне және осындай мәтіндер типтерінің семантикалық көрінісінің толық құрылуына да байланысты екенін айта кету қажет.

Қолданбалы лингвистиканың дамуына, яғни ғылыми жетістіктерге ие болуына тіл білімі теориясының дамуы да өз әсерін тигізбей қоймады. Мысалы, ХХ ғ. 20–30-жылдары практикалық қажеттіктен туындаған ғылыми-техникалық терминологияны бірізді және тұрақты ету жолдары жаңа лингвистикалық пән – “терминтану” (“терминоведение”) саласын өмірге әкеледі. Сол сияқты, Кеңес одағы халықтары тілінің әліпбиі мен жазуын құруға байланысты 30–40-жылдары жүргізілген кең көлемді әрі орасан күрделі, әрі зор тәжірибелік мәні бар жұмыстар тілдерді синхронды сипаттаудың әдістерін жетілдіруге себепші болды (ынталандырды) деуге болады.

Соңғы 20–30 жылда пайда болған қолданбалы лингвистиканың жаңа аспектілерінің бәрі бір ғана ортақ проблемаға тәуелді екенін байқатты, ол – тілдің жазба не сөйлеу түрлерін автоматты өңдеу мәселесі еді. Мұндай проблемамен айналысу тілді талдау мен сипаттаудың жаңа әдіс-тәсілдерінің дамуына және тіл табиғаты мен тіл білімі құрылымына деген жаңаша көзқарастың пайда болуына мүмкіндік туғызды.

ХХ ғасырдың 50–70-жылдарында тілдік материалдарды автоматты түрде өңдеуге қатысты күрделі проблеманың шешім табуы қолданбалы лингвистиканың әрі қарай дамуына айтарлықтай әсер етті.

Қолданбалы лингвистиканың жаңа аспектілері теориялық тіл біліміне жаңадан ғана ене бастаған математикалық әдістерді, әсіресе, теоретика-жиындық, формальды-логикалық, статистика-ықтималдық әдістердің қолдануын барынша жеделдетті.

Классикалық тіл білімінің, классикалық логиканың, психологияның, семантика мен математиканың қиылысуынан барып тілдегі модельдеу әдісінің өзінше бөлек ғылым ретінде көрініс табуының нәтижесінде қазіргі “құрылымдық лингвистика” деп аталып жүрген ғылыми сала өмірге келгені белгілі. Сол сияқты енді тіл біліміндегі дербес тұрған теориялық бағыт ретінде “тілдің теориялық модельдері” атты жаңа тарау пайда болды.

Лингвистика мен математика салаларының аралығынан туындаған – “математикалық лингвистика” атты жаңа пән өмірге келді. Бұл пәннің ХХ ғ. 50–60-жылдары қалыптасуы бұрыннан да тіл білімінің барлық ішкі дамуының өзінен де сезіле бастаған болатын. Мәселен, бұл жайт тілдің құрылымына көңіл аударудан, тілдік жүйе ретінде қарастырудан және оның микрожүйелерден тұру күрделілігін жете түсінуден, лингвистикалық нысандар мен олардың атрибуттары аралығындағы қатынастарды танып-білуден байқалып, математикалық лингвистика пәнінің дамуына объективті жағдай туғызды деуге болады.

Атап айтқанда, қазіргі теориялық тіл білімінің дамуына ең көп әсер еткен – құрылымдық лингвистика мен математикалық лингвистика пәндері. Бұндай әсер көбінде грамматиканың синтаксис және семантика салаларында айтарлықтай көрініс тапты.

Құрылымдық синтаксисте мынадай формальды екі синтаксистік модельдер жүйелі түрде жете зерттелді. Олар – тікелей құрастырушылар моделі мен өзара тәуелді модельдер деп аталады. Бұл аталған екі модель де машиналық аударма жасауда,

автоматтанған синтаксистік талдау мен мәтінге жасалатын басқа да автоматты өңдеулерде кең түрде пайдаланылады.

Семантика тілдің барлық деңгейлеріне ортақ болғандықтан, тіл білімінде оны бөлек деңгей ретінде қарастырмайды. Семантикада, біріншіден, лингвистиканың өзіне ғана тән әдістің әсері байқалса, екіншіден, логикалық семантика мен теориялық классификациялау әдістерінің де әсері орын алады. Синтаксистік семантикада сан жағынан көбірек зерттелген жайттар – сөйлемнің семантикалық құрылымы жайлы тұжырымдар (концепциялар).

Қоғамдық ғылымдарда ғылыми зерттеулерді автоматтандыру, ең алдымен, құжаттық және фактографиялық деректердің ауқымды қорын құру барысында қажет болды. Себебі, мұндай автоматтандыру типі ең алдымен қоғамдық ғылымдар мамандарының өз қалауынан туындап отыр. Мәселен, ғылым жолындағы қандай анықтамалар түрі қоғамдық саладағы қызметкерге қажеттірек? Біріншіден, мұндай сұраныс кітапханалардағы библиографиялық істе: белгілі бір мәселеге қатысты барлық отандық не шетелдік әдебиеттерді іздестіріп, тауып алу сияқты мәселелер. Мұндай материалдарға жататындар: кітаптар, мақалалар, конференция тезистері, хроникалық шағын мақалалар және т.б.

Екінші кезеңде, жиналған әдебиеттер ішінен қажетті деректерді жинастыру, оларды қажетті белгілеріне қарай сұрыптау, топтастыру және т.б. жұмыстар орындалуы керек.

Көптеген қоғамдық ғылымдарға қатысты мәселелер, әсіресе, тіл білімінде, этнографияда, антропологияда, тарихта карталық мәліметтерге тікелей қатысты болып келеді. Сондықтан, ЭЕМ-ге ондай деректерді енгізу мен қағаз бетіне шығару өте-өте өзекті деп саналады.

Қоғамдық ғылымдар бойынша ғылыми-зерттеу жұмысын автоматтандыруды қажет ететін мәселелер:

- а) әр тілдер бойынша белгілі тақырыпқа қатысты әдебиеттерді іздеп, табу;
- ә) жазба ескерткіштер мен мәтіндік материалдар қоры бойынша таңдама жұмыстарын атқару;
- б) табылған материалдар арқылы қажетті амал-әрекеттерді іске асыру;
- в) материалды жан-жақты сұрыптау;
- г) алдын ала белгілі өлшемдер бойынша реестрлер, каталогтар сипаттамаларын жинақтау;
- д) статистикалық, картографиялық, сұрыптау теориясы мен жүйелік талдау әдістерін қолдану;
- е) берілістерді (деректерді) сызба, сурет, карта түрінде көрсету.

Сонымен, ғылыми жұмысты автоматтандыру процесінде құрастырылған жүйе негізінде тұғынушы маманға қажетті тарихи деңгейде зерттеу нысанына қатысты барынша толық энциклопедиялық білім және библиографиялық мәліметтер берілуі қажет деп санаймыз.

Жоғарыда сөз болған мәселелерді қорыта келе айтатын болсақ, қолданбалы лингвистика, ең алдымен, кешенді пән. Себебі, ол философия, психология, физиология, математика, логика, әлеуметтану, информатика салаларымен үнемі тығыз қатынаста болады.

Жаңаша мағынадағы қолданбалы лингвистиканы жеке қарастыратын болсақ, ол көптармақты ғылыми сала болып табылады. Ғылыми әдебиеттерде оны “компьютерлік лингвистика”, “инженерлік лингвистика”, “автоматты лингвистика”, “есептеу лингвистикасы”, “квантитативті лингвистика”, “статистикалық

лингвистика” деген жарыспалы терминдермен атап жүр. Әрине, бұл атауларға сәйкес салалардың өздеріне тән бағыт-бағдары ажыратылады, алайда бәріне тән ортақ мақсат — адам баласының қарым-қатынасы үшін қажетті табиғи тіл қызметінің ең ыңғайлы, ең тиімді жолдарын іздестіру.

Қолданбалы лингвистика салаларының барлық тілдерде, соның ішінде қазақ тілінің жазба түріне қатысты зерттеулерде ең көп қолданыс тапқан тармағы — статистикалық лингвистика. Дәстүрлі әдіспен тілді зерттейтін тілші-ғалымдардың бәрі бірдей статистикалық әдіс-тәсілдерді қолдай бермейді, бірақ өз ізденістерінде осы саланың *көп, аз, мол, жиі, сирек, өнімді, өнімсіз* тәрізді терминдерін барынша пайдаланады және олардың статистикалық ғылым саласына қатыстылығын көп жағдайда аңғара бермейді.

Статистикалық лингвистика тілді зерттеудің тек әдісі не тәсілі болып қана қалмай, қазіргі кездегі тілтануда өзінше бір бөлек ғылыми пән дәрежесіне көтеріліп отыр. Бұл пән тілдік бірліктердің сапалық және сандық (мөлшерлік) мөлiметтерiн тілдің табиғатына сай қарастыруды жөн санайды.

Статистикалық лингвистика, негізінен, математикалық статистика әдістемесіне сүйенеді. Зерттеу барысында ондай әдістемені қолдану қатынас құралы ретіндегі “тілді” белгілі бір жүйе деп қарастырудан туындайды. Шынында да, “тіл” дегеніміз ақиқат болмыста өмір сүретін көп өлшемді және белгілі бір ішкі заңдылықтардың негізінде топталған (реттелген) табиғи құбылыс деуге болады. Ал тілдік жүйе математикалық статистика заңдылықтарына бағыну үшін ондағы бірліктер тобы қайталанып отыратын және кездейсоқтық сипатта болатын тілдік элементтерден тұруы қажет. Бұл жерде айта кететін жайт, тіл қызметі кезінде (яғни тілдік қарым-қатынас кезінде) дыбыстар, сөздер, сөз тіркестері, сөйлемдер және одан да үлкен бөліктер қайталанбайтын сипатта болса, адамдар арасындағы тілдік қатынас бұзылып, ақпарат таратушы мен қабылдаушы арасында түсінбестік пайда болған болар еді. Сондықтан да тілдік бірліктердің қайталануы олардың тілімізге тән болуынан, олардың қолдану сипатына *жиі, сирек* деген ұғымдардың да тән екендігінен туындайды.

Табиғи ортадағы элементтер арасындағы қайталану сипатының негізінде математикалық статистика мен ықтималдық теория атты ғылым салаларында тілдік бірліктердің статистикалық заңдылықтары қалыптасады. Зерттеуге алынған бірліктің (единица, элемент) ондай заңдылыққа бағыну не бағынбауын айқындау үшін алынатын ортаның аумағы (мәселен мәтін көлемі) барынша мол болғаны жөн. Мысалы, кітаптың бір бетінде кездесетін сөздер мен оның жүз бетінде кездесетін сөздердің қайталану сипаты әртүрлі болатындығына ешкімнің күмәні жоқ. Ал заңдылықты айқындау үшін, кітаптың бір бетіндегі мөлiметтен гөрі оның көптеген беттерінен алынған мөлiметтердің шамасы шындыққа жақын болатындығы белгілі.

Тілге тән заңдылықтарды статистика тәсілімен анықтау негіздемесі мынада:

- 1) сандық (мөлшерлік) құбылыс тіл табиғатына әуелден-ақ тән болуы;
- 2) тіл құрылымындағы сандық және сапалық сипаттардың өзара байланыста болуы;
- 3) тілдің әртүрлі бірліктері сөйлеу ағымында статистикалық заңдылықтардың ең болмағанда біреуіне бағынуы тиіс деп ұйғару керектігі.

Әдістердің индуктивті, дедуктивті болып ажыратылатынын ескере отырып, негізінен алғанда, статистикалық әдіс индуктивті әдіс арқылы, ал тілдің әртүрлі модельдерін жасау үшін қолданылатын логика-математикалық әдістер — дедуктивті әдіс арқылы жүзеге асады.