

Л 2007
8661_к



КОМПЬЮТЕРЛІК
ЛИНГВІСТИКАҒА
КІРІСПЕ

А л м а т ы

А. Қ. Жұбанов

**КОМПЬЮТЕРЛІК
ЛИНГВИСТИКАҒА
КІРІСПЕ**

Оқу құралы

Алматы
«Қазақ университеті»
2007

ББК 81.2
Ж 83

*Баспаға Әл-Фараби атындағы Қазақ ұлттық
университеті филология факультетінің Ғылыми кеңесі
және Редакциялық-баспа кеңесі ұсынған*

Пікір жазғандар:

филология ғылымдарының докторы, профессор **Ф.Оразбаева;**
филология ғылымдарының докторы, профессор **Г. Б. Мәдиева**

Жұбанов А.Қ.
Ж 83 Компьютерлік лингвистикаға кіріспе: Оқу құралы. –
Алматы: Қазақ университеті, 2007. – 204 б.

ISBN 9965-30-284-7

Оқу құралында басты назар филология саласында электрондық есептеу машиналарын қолданудың гуманитарлық аспектілеріне аударылады. Еңбекті филологиялық пәндерді оқытуда және жоғары оқу орындарындағы әртүрлі арнайы курстар бойынша дәріс беруде пайдалануға болады. Қолданбалы лингвистиканың дәстүрлі және жаңа салалары: жасанды интеллект, машиналық аударма, ақпарат іздестіру жүйелері, қазақ тілінің компьютерлік қорын құрастыру жайттары және т.б. қарастырылған.

Кітап университеттер мен педагогикалық институттардың студенттері мен магистранттарына, математикалық және қолданбалы лингвистика мамандарына, компьютер мүмкіндіктерін қоғамдық ғылымдар салаларына қолдану мәселелерімен айналысатын ғалымдарға арналған.

Ж 4603000000-089 088-07
460(05)-07

ББК 81.2

ISBN 9965-30-284-7

© Жұбанов А.Қ., 2007
© Әл-Фараби атындағы ҚазҰУ, 2007



АЛҒЫ СӨЗ

Табиғи тілдегі ақпараттардың басым көпшілігі сөйлеу тілі түрінде немесе жазба мәтін түрінде кездесетін болғандықтан, табиғи тілдік ақпаратты компьютер арқылы өңдеудің маңыздылығы артуда. Осы айтылғандарға сәйкес мәселелердің қолданбалы және теориялық лингвистикаға, қолданбалы филологиялық зерттеу аясына да қатысы бар. Оның қолданбалы филология саласына қатыстылығы – мәтінге түсініктеме беретін ғылымдар қатарына жататындығынан. Сонымен бірге қазіргі кездегі бұқаралық қатынас құралдары мен іс прозасының барынша даму сатысында оның маңыздылығы бұрынғыдан да арта түсуде.

Ақпараттың әрбір түрі өз мазмұны арқылы сипатталады да, бірақ олардың бәрі бірдей машиналық (компьютерлік) көрініске қатысты жалпы заңдылыққа бағынады. Сондықтан информатика ғылымының негізгі міндеті ақпараттың барлық түрін қазіргі техникалық құралдар (компьютерлер) арқылы дайындау, өзгерістер енгізу, ұзақ мерзімге сақтау, тұтынушы қауымға тарату мен оны қолданудың заңдылықтарын зерттеу.

Ғылым ретіндегі информатика саласының кең мағынадағы түсінігі бойынша, ол әлемнің (ғаламның) ақпараттық моделін зерттеу мен құрастыру мәселесімен айналысатын ғылым саласына айналып отыр. Информатика мәселелерінің негіздері тілдік және мәтіндік бірліктердің атқаратын рөлімен жанасып жатады. Бұл бірліктер ең кіші құрастырылымдардан бастап, олардың ең ірілерін қоса есептегенде барлық тілдік деңгейлерді

қамтиды деуге болады. Тілдің әр деңгейіне және оның толық пішініне немесе ішкі жүйелеріне түсініктеме бере алатын әмбебап сипаттағы лингвистикалық теория әлі де жоқ және ондай теорияның болу-болмауының өзі де күмән туғызады. Сондықтан модельдеу мен зерттеудің нысаны ретінде тіл қызметінің әр аспектілері жеке қарастырылуы керек.

Енді компьютерлік лингвистика саласына тоқталатын болсақ, оны да тіл ғылымының осындай жеке бөліктерін зерттейтін ғылым саласы деуге болады. Компьютерлік лингвистика – информатиканың лингвистикалық негіздерін және тілдік байланыстардың барлық аспектілерін, ойлау мен ақиқаттықты модельдеу әрекеттерін компьютерлік программалар арқылы тікелей зерттеп танумен айналысатын ғылым саласы. Сол сияқты, компьютерлік тіл білімі тілдік бірліктердің сөзден кіші, сөзге тең, сөзден үлкен түрлерін, яғни сөз тіркестерін, сөйлемдерді (айтылымдарды), толық мәтін мәселелерін және тілдік әрекеттерді модельдеу проблемаларын зерттейді. Мысалы, соңғы айтылғанға қатысты, мәтін ішіндегі мән-мағынаны анықтау мен бір тілден екінші тілге аудару мәселелерін атауға болар еді. Лингвистикалық автоматтардың мүмкіндіктерінің күннен-күнге артуына байланысты компьютерлік лингвистиканың шешетін мәселелерінің де аясы кеңейіп, тереңдей түсуде. Оның шеңберінде жаңа әдістер мен теориялар тәжірибе жүзінде зерттеліп, сыннан өтуде.

Біз ұсынып отырған оқу құралында «қолданбалы лингвистика», «инженерлік лингвистика», «информатиканың лингвистикалық негіздері», «компьютерлік лингвистика» және т.б. терминдер қолданылады. Олардың әрбіреуіне жеке тоқталмай-ақ, бәрінің де «қолданбалы лингвистика» пәнінің аясына сыйысып кететінін, оның кең мағыналы екендігін айтуға болады. Қолданбалы лингвистика термині компьютерлік лингвистика мен математикалық лингвистиканы ғана біріктіріп қоймайды, сонымен бірге лингводидактика мен одан да басқа ғылымдар салаларын қамтиды. Жоғарыда аталған терминдер ішіндегі «информатиканың лингвистикалық негіздері» атты терминнің ең тар мағынаға ие екенін ескеру қажет. Оны, шын мағынасында, информатиканың компьютерлік және ақпараттық ғылымдар салаларымен тығыз қатыстылығынан туындайтын

лингвистикалық мәселелердің жиынтығы деп те қарастыруға болады. «Компьютерлік лингвистика» терминін аталған екі ұғымның аралығындағы мағынаға ие деуге болады. Ол «информатиканың лингвистикалық негіздеріне» тән мағынадан кең түрде қолданылады. Дәлірек айтқанда, мысалы, ол лингводидактикаға көбірек қатысы бар оқыту-үйрету жүйесінің лингвистикалық мәселелерімен де айналысады. Бірақ «компьютерлік лингвистика» терминінің «қолданбалы лингвистика» терминіне қарағанда мағынасы тар.

Қазіргі кездегі компьютерлік лингвистика саласының кейбір қолданыс табу аясын осы оқу құралында қарастыруды мақсат еттік.

Компьютерлік лингвистика ғылымының құрамы мен құрылымы жайлы пікірлер әлі де болса тұрақтала қойған жоқ. Бірақ мұндай жағдай әр ғылым саласының да бастапқы кезеңіне тән екені белгілі. Сондықтан компьютерлік лингвистика туралы сөз болғанда оны қолданбалы лингвистиканың компьютерге, компьютерлік программалар жазуға, жасанды интеллектіге қатысты мәселелері деп түсінген жөн.

Информатиканың лингвистикалық негіздерін әртүрлі түсінікте құрастыруға болады. Мәселен, Ю.Н.Марчук «Основы компьютерной лингвистики»¹ атты оқу құралын жазуда өз алдына компьютерлік лингвистикаға байланысты «информатиканы» тілдік деректермен қамтамасыз етудің нәтижелерін жинақтау мен даму болашағын белгілеу мақсаттарын қойған.

Ал қолданбалы тіл білімі бойынша, ең алғаш 1996 жылы С.-Петербургте жарық көрген «Прикладное языкознание» атты оқулықтың² құрылымы мен мақсат-міндеттері жоғарыда аталған еңбектен басқаша. Онда тіл білімінің қолданбалы саласының негізгі мәселелері (әдістемеліктен жалпығылымдыққа дейін) дербес қарастырылған. Басқаша айтқанда, оқулыққа енген материалдар олардың атқаратын қызметіне қарай топтастырылып, жеке-жеке тақырыптар бойынша көрініс тапқан. Мысалы,

¹ Основы компьютерной лингвистики: Учебное пособие. М.: Изд-во МПУ «Народный учитель», 2000. 226 с.

² Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. 528 с.

“Орфография”, “Автоматизированные обучающие системы”, “Психолингвистика”, “Учебная лексикография”, “Машинный перевод” және т.б.

Аталған оқулықтың оқыту процесіндегі теориялық және әдістемелік артықшылықтарын ескере отырып, біз ұсынып отырған “Компьютерлік лингвистикаға кіріспе” атты оқу құралында да сол құрылымдық және мазмұндық принциптер ұсталынды. Сонымен бірге, ол оқулықтың компьютерлік лингвистикаға қатысты тақырыптары автордың ұғынуынша қазақшаға аударылып, әл-Фараби атындағы Қазақ ұлттық университеті филология факультетінде “Қолданбалы лингвистика” мамандығы бойынша оқитын студенттерге (2001 жылдан бері) және Абай атындағы Қазақтың ұлттық педагогикалық университеті қазақ филологиясы факультетінің магистранттарына дәріс беруде пайдаланылғанын атауға болады. Оқылған теориялық материалдар қажетті жүйеге келтіріліп, ұсынылып отырған оқу құралына енді. Оның құрылымына келетін болсақ, ол алғы сөзден, кіріспеден, компьютерлік лингвистика саласының маңызды мәселелеріне қатысты 11 күрделі тақырыптан және қорытындыдан тұрады. Әрбір тақырыпқа қатысты пайдаланылған әдебиеттер тізімі әрбір тақырыпқа қатысты мақаланың соңында берілді.

Әрине, бұл тақырыптардың бәрі бірдей толық дәрежеде баяндалды деуден аулақпыз. Еңбектің негізгі мақсаты – жоғары оқу орындары студенттеріне компьютерлік лингвистика ғылымына кіріспе ретінде мәлімет беретін қазақ тіліндегі оқу құралын ұсыну. Ескертпе түрінде айта кететін жайт, оқу құралы мәтінінде кейбір орыс тіліндегі және орыс тілі арқылы енген салалық терминдер автордың өз ұйғаруымен қазақ тіліне аударылып берілді.

КІРІСПЕ

«Компьютерлік лингвистика» термині (computational linguistics), негізінен, тіл қызметін модельдеуге қатысты пайдаланылатын компьютерлік бағдарламалар мен тілдік деректерді өңдеу, сол үшін қажетті компьютерлік технологияны жинақтау және компьютердің араласуына қатысты басқа да тіл зерттеу жұмыстарымен байланысты қолданылады. Кең мағынада алғанда, тіл білімі аясына қатысты теориялық және практикалық мәселелердің шешімін табуға байланысты компьютердің араласу жағдайларының бәрін бірдей компьютерлік лингвистика саласына жатқызуға болады.

Адам баласының ақпараттық ортада өмір сүретіні белгілі. Ал бұл ортада қабылдаушының жауапты ақпарат бойынша дұрыс шешімге келуіне жағдай жасайтын ақпараттық технологияның түрлері көп. Олар: автоматтандырылған ақпараттық жүйе, бұқаралық коммуникация құралдарына жататын теле-радио байланыс жүйелері, компьютерлік (машиналық) аударма жүйесі, диалогтік жүйе және сұрақ-жауап, т.б. жүйелер. Мұндай жүйелердің саны мен сапасы күннен-күнге артуда. Ақпараттық қоғамдағы тілдік практика мен қазіргі кезеңдегі тілді зерттеу теориясының қолданылу аясы да өсіп келеді.

Тілді зерттеудегі компьютерлік лингвистика саласы қолданбалы тіл білімі атқаратын қызметтердің бір тармағы ғана деуге болады, себебі «қолданбалы тіл білімі» ұғымы көптармақты бола келе, күнделікті өмірде дәстүрлі және жаңаша

мағынада кездесіп жүр. Мәселен, бұл ұғым дәстүрлі түсінік бойынша, тіл қызметі мәселелерінің практикалық әдістерін жасауды, яғни жазу мәселесін дамытуды, ауызекі сөйлеудің транскрипциялық жүйесін құруды, шетел сөздерінің транскрипция жүйесін жасауды, стенография тәсілдерін дамытуды, арнайы лингвистикалық сөздіктер шығаруды, тарихи ескерткіштердегі таңба-жазуларды айқындауды, ғылыми-техникалық терминологияны қалыптастыруды т.б. іс-шараларды қамтиды.

Қолданбалы тіл білімінің жоғарыда сөз болған дәстүрлі саласы қазақ тілінде ойдағыдай дамып, түркі тілдері ішінде көш бастап келеді деуге болады. Ал оның компьютерге байланысты жаңа саласы — «компьютерлік қазақ тіл білімі» әлде де кешеуілдеп дамуда.

Қолданбалы тіл білімінің «статистикалық лингвистика» (квантитативті лингвистика) тарауы бойынша ғылыми ізденістер ғалым Қалдыбай Бектаевтың атымен және оның ғылыми мектебімен тығыз байланысты. Қазақ тілінің әртүрлі стильдерінің мәтіндері бойынша, алғашында электронды-есептеу машинасының (ЭЕМ), кейіннен компьютерлердің көмегімен алынған жиілік сөздіктер (әліпбилі жиілік, жиілік, кері әліпбилі) тілдің құрылымдық жағының статистикасы жайында көптеген мәліметтер алуға мүмкіндік туғызды. Осылайша, қазақ тіліндегі бай статистикалық мәліметтер негізінде бірнеше кандидаттық және докторлық диссертациялардың қорғалғаны көпшілікке мәлім. Бүгінгі таңда А.Байтұрсынұлы атындағы Тіл білімі институтының «компьютерлік лингвистика» бағыты бойынша қол жеткен ғылыми нәтижелері айтарлықтай деуге болады. Осылардың тек негізгілері деп мынадай ғылыми жұмыстарды атауға болады:

— көркем әдебиет, көсемсөз, ғылыми-көпшілік әдебиет пен оқулықтар мәтіндерінің, алғашында ЭЕМ-ның, кейіннен компьютердің жадына көптеп енгізіліп, олар бойынша арнайы компьютерлік бағдарламалар құрастырылуының негізінде әртүрлі жиілік сөздіктердің алынуы;

— М.Әуезовтің 20 томдық шығармалар жинағының компьютер жадына енгізілуі және соның негізінде жазушы тілінің жиілік сөздіктерінің жарық көруі;

– осы тәрізді жиілік сөздіктердің негізінде қазақ сөздерінің әр стильде және әр көлемді мәтіндерде қайталануының қосынды жиілігінің, сол мәтіндерді қамту статистикасының анықталуы;

– сөзтұлғалардың құрылымдық статистикасы мен мәтіндегі пайыздық салмақтарының әр сөз таптарына қатысты айқындалуы;

– қазақ тіліндегі негізгі сөз таптарының мәтін бойында кездесулерінің математикалық статистика мен ықтималдық теория заңдылықтарына бағыну-бағынбау жағдайларының зерттелуі;

– қазақ тілінің «ТІЛ – ҚАЗЫНА» атты автоматтандырылған компьютерлік картотекалық қорының іске қосылуы;

– қазақ тіліндегі мәтін мазмұнын оның тұрпатына қарай ашудың (формалдаудың) негізгі принциптерінің айқындалуы;

– Тіл білімі институтында жоспарлы тақырыптар бойынша түзіліп жатқан түсіндірме, екі тілдік, бір тілдік сөздіктер мәтіндерінің компьютер жадында «сөздіктер қорын» құрап, тұрақты сақталуы.

Осы аталған іс-шаралардың қатарын, компьютердің араласуымен орындалған жұмыстардың және компьютерлік тіл білімі саласына қатысты басқа да жұмыстарды атап өтуге болады.

Компьютерлік лингвистика бағыты түркі тілдері үшін тіл зерттеудің жаңа саласы деп есептелгенімен, үнді-еуропа тілдерінде дәстүрлі болып, күнделікті зерттеу мәселелеріне айналған компьютерлік лингвистиканың кейбір жайттарына тіл зерттеудегі жаңаша көзқарасқа қызығушы қауымға арнап, арнайы сөз етуді жөн көрдік.

Өткен ғасырдың 50-жылдарынан бастап «қолданбалы лингвистика» термині жаңа мағыналарға ие бола бастады. Ол күнделікті өмірімізге компьютерлік технологияның және автоматтанған басқару жүйесінің араласуы арқылы мәтінді автоматты түрде өңдеуге қатысты көптеген қолданбалы бағыттағы тілдік бірліктерді талдау (анализ) мен жинақтау (синтез) негізінде тілдің танылуы мен талдану заңдылықтарын белгілеу қажеттігінен туындайды. Орыс тілінде жарық көрген әдебиеттерде «қолданбалы лингвистика», «есептік лингвистика», «инженерлік лингвистика», «автоматты лингвистика», «матема-

тикалық лингвистика», «статистикалық лингвистика» терминдері бірінің орнына бірі қолданыла бастады. Әрине, бұл аталған терминдерді ортақтастыру дұрыс шешім деуге болмайды, себебі олардың қолданбалы тіл саласындағы мақсаты мен міндетінде және әдіс-тәсілдерінде өзіндік ерекшеліктері бар. Тілдің қолданбалы шеңберінің кеңею сипатын оның әртүрлі ғылымдар салаларымен түйісу мүмкіншілігінен деп түсіну керек (*Городецкий, 23*). Тіл қызметінің пәнаралық сабақтастық қасиеті тілдік теорияға тек қана оң әсерін тигізумен қатар, қазіргі тіл білімінің тұжырымдамалық бағытын (концептуальдық аппаратын) жаңартуға да мүмкіндік туғызды.

Қолданбалы лингвистика саласының кең мағыналылығын — тілдің құрылымы мен қызметіне қатысты ғылыми мағлұматтарды тілге қатыссыз пәндер саласында, адам баласының практикалық мұқтаждығынан және мұндай қызметтің теориялық негіздемесі ретінде қолданылуынан туындайтынын көруге болады.

Қолданбалы лингвистиканы функционалдық көзқарас тұрғысынан қарастырсақ, оны тіл қызметінің ең қолайлы әдіс-тәсілдерін ұсынатын ғылыми пән деп те түсінуге болады. Тілдің коммуникативті функциясының қолайлы жақтарын ұсынатын пәндер — аударма теориясы, машиналық (компьютерлік) аударма, ана тілі мен шет тілдерін оқытудағы ақпараттық-ізвестіру жүйесінің теориясы мен практикасы, ақпараттық жасанды тіл жүйесін жасаудағы таңбалау теориясы.

Тілдің коммуникативті функциясының бір көрінісі ретіндегі әлеуметтік қызметі әлеуметтік тіл білімінде, тілдік жоспарлау мен тілдік саясатта, орфография мен орфоэпияда, әсерлік теория мен саяси лингвистикада көрініс табатыны белгілі. Сол сияқты тілдің когнитивті қызметінің ең қолайлы жақтары компьютерлік тіл білімінде, психолінгвистика мен афазиологияда және квантитативтік лингвистика салаларында көрініс тауып жүр (*Баранов, 8*).

Компьютерлік лингвистика информатиканың тілдік негіздемесін қарастыруымен қатар «тіл» мен «ойлау» арасындағы байланысты және одан туындайтын тікелей болмысты компьютер арқылы модельдеуге қатысты мәселелермен шұғылданады. Компьютерлік лингвистиканы тілдік бірліктерге

қатысты проблемалар да қызықтырады. Ондай бірліктер сөзден кіші, үлкен және оған тең де болуы мүмкін. Бұл жердегі сөзден үлкен деп отырғандарымыз — сөз тіркестері, сөйлемдер (айтылымдар) және толық мәтіндер. Аталып отырған салада тілді зерттеу кезінде қолданылатын амал-тәсілдерді модельдеу жолдары да қарастырылады. Мәселен, мәтін мазмұнын ашу жолдарын немесе аудармашының бір тілдегі мәтінді екінші тілге аудару жолдарын модельдеу мәселелері бұған мысал бола алады.

Сөйлеу мен ойлау процестерінің арасындағы байланыстардың аса күрделі философиялық мәселе екені мәлім. Біз олардың түрлерін көптеген айтылымдардағы, сөйлемдер мен мәтіндердегі тілдік мәліметтерді тұрпатына қарай (формалды) сипаттауда ұтымды пайдалануға болады демекпіз.

Адам баласының «ойлау» қабілетін модельдеу жағдаятында ғалымдардың назары көбіне лингвистика мен логика арасындағы байланысқа ауысып, тұжырымдау процесі ойлаудың анықтаушы және шешуші сипаты болып саналады. Қазіргі тіл білімінде тіл мәселесіне байланысты кейбір жайттарды логика саласының мамандары қарастырып, ал кейбір логикалық тұрпаттар (формалар) сырын ашу жағы тілшілерге жүктелуде. Бұл түсінікті де. Себебі, грамматикалық (семантикалық) және логикалық тұрпаттар әр көзқараспен қаралатын бір ғана нысанның түрлері болып келеді. Табиғи тіл адам баласындағдылы ақиқат шындығынан уақытша ажыратып, тілдік тұлғалардың семантикалық мағынасымен анықталатын «ойлау» формаларымен әрекет жасайтын жаңа деңгейге ауысуына жағдай жасайды.

Бұл айтылғандардың барлығы да компьютерлік лингвистика үшін аса маңызды мәселелер. Себебі, адам баласының ойлау қабілетін тұрпатына қарай (формалды түрде), әртүрлі дәлдікпен қайталаудың мүмкіншілігі туады. Ал бұл жағдай математикалық модельдеу мен машиналық логиканың көмегімен іске асатын «ойлау» қабілеті бар «жасанды интеллект» жасап шығаруға мүмкіндік туғызады. Бұл жердегі «жасанды интеллект» атты жүйені — іс жүзінде «ойлау» қабілеті бар және «шығармашылыққа» қатысты мәселелердің шешімін

таба алатын компьютерлік бағдарлама (программа) деп түсінген жөн.

«Жасанды интеллекті» іске қосу үшін ең алдымен адам баласының белгілі бір жағдайда шешім қабылдаудағы немесе жауапты мәселеге шешім іздеу кезіндегі ойлау процесінің сырын зерттеп тану қажет. Әсіресе, адамның ойлау процесіндегі көптеген сатылардың ішінен ең маңыздыларын бөліп алуды мақсат еткен жөн. Бұл жолда күрделі ойлау қызметінің құрылымы жағына мән бере отырып, шығармашылықтың құпиясын қарапайым сатылы бөліктерге бөлшектеп, компьютерлік бағдарламаның алгоритмін құрудың, болашақ «ойлау» қабілеті бар «жасанды интеллект» жасап шығарудың алғашқы кірпішін қалау жолдары зерттеледі.

Ми қызметін дәлме-дәл анықтау мүмкін бола бермейді, себебі адамға тән ақыл-ой – өте күрделі әлеуметтік-биологиялық функция, сондықтан да оның құпиясын ашу көптеген зерттеулерді қажет етеді.

«Жасанды интеллектіге» қойылатын негізгі талап ми қызметінің моделін «ойлау құралы» деп ұйғарудан туындайды. Тіл қызметіне қатысты модельдеудің негізгілері болып есептелетін – аударма мәселесі. Бұл тілдік құрылымның морфологиясынан синтаксисіне дейінгі барлық деңгейін қамтиды. Сондықтан да компьютерлік аударма тіл қызметінің аса күрделі функцияларының бірі болғандықтан, оны модельдеу де оңайға түспейді. Қазіргі кезде көптеген еуропа тілдерінен орыс тіліне (және керісінше) компьютерлік аударма жасау мәселесі біршама шешілді. Ал осы сияқты аудармаларды түркі тілдері бойынша, оның ішінде қазақ тіліне қатысты жүзеге асыру – егеменді еліміздің кезек күттірмейтін мәселелерінің бірі деуге болады. Қазақ тілінің компьютерлік саласы жаңа ғана дамып келеді және оның ойдағыдай дәрежеге жетуіне барлық мүмкіндік бар. Себебі, компьютерлік лингвистика саласы жеке ғылыми бағыт ретінде өткен ғасырдың 60-жылдарында ғана қалыптасты. Бұл сала бойынша жарық көрген ғылыми жұмыстар жеткілікті. Мәселен, АҚШ-та әр тоқсан сайын «Компьютерлік лингвистика» атты ғылыми журнал жарық көреді. Осымен байланысты компьютерлік лингвистика мәселесі бойынша айналысатын аймақтық құрылымдағы әлемдік «Қауымдастықтың» және әрбір

екі жыл сайын осы сала бойынша өткізілетін «КОЛИНГ» атты халықаралық конференцияның жұмыстары да айтарлықтай жемісті деуге болады (*Баранов, 14*).

«Компьютерлік лингвистиканың» айналысатын мәселелерінің ең маңыздыларының бірі – табиғи тілдің мәтін морфологиясын автоматты түрде талдау. Егер, дәстүрлі түсінік бойынша, «сөз морфологиясы» деген ұғым сөздің тұлғалық құрылымын, яғни сөз түбірін және оған жалғанатын қосымшалардың (жұрнақ, жалғау) түрін анықтайтын болса, компьютерлік тіл біліміндегі «морфология» терминінен басқаша ұғым туындайды. Дәлірек айтсақ, ол сөздің мәтін бойындағы сыртқы тұрпатына қарай, тілдік құрылымның әрқилы деңгейлері бойынша тілдік мәліметтер алудың мүмкіндігін білдіреді. Алғашында «морфологиялық талдау» ұғымы айтылған мағынада машиналық аударма саласында пайда болды (*Василевский*).

«Морфологиялық талдаудың» жаңаша түсінігі бойынша алынатын ақпарат біздің дәстүрлі ұғымымыздағы морфологияға қатысы жоқ көптеген амалдардан тұруы мүмкін. Осының салдарынан компьютерлік тіл біліміндегі «морфологиялық талдау» ұғымы – амалдар ұғымы (операционное понятие), яғни сөздің сыртқы пішіні бойынша тілдік ақпараттарды танып-білу үшін жүргізілетін іс-әрекет деп ұққан жөн сияқты. Басқаша айтсақ, егерде дәстүрлі тіл білімі ұғымы бойынша «морфологиялық талдау» кезінде «нені талдаймыз?» деп сұрақ қоятын болсақ, компьютерлік тіл білімінде «қалай талдаймыз?» деген сұрақ қойылуы керек, яғни ол сөз тұлғасы бойынша қажетті тілдік ақпараттарды қай жолмен алуға болатындығымен айналысады.

Машиналық (компьютерлік) аударманың алғашқы тәжірибелерінде мұндай амалдардың саны көптеп кездесетін, ал бүгінгі таңда орыс және еуропа тілдері бойынша бұл мәселе біршама шешілді деуге болады (*Лингвистические...*). Ал қазақ тілі мен басқа да түркі тілдері бойынша мұндай ізденістер айтарлықтай қолға алынбай жүр.

Аталған түсініктегі «морфологиялық талдау» бірнеше бағытта қарастырылуда. Олардың біріншісі – сөзтұлғаны оның негізі мен оған жалғануы мүмкін болатын қосымшаларға ажырататын классикалық талдау негізінде модельдеу.

Екінші бағыт сөзтұлғалардың соңғы әріптерінің тіркесімдік заңдылықтарының жиілік сөздіктердегі статистикалық мәліметтеріне сүйенеді.

Үшінші бағыт — соңғы кездердегі ізденістердің нәтижесі. Бұл бағыт бойынша теңдеулердің ашық жүйелерінің пішіні ретінде морфологияның әмбебап математикалық моделі жасалады. Модельдің есептегіш мүмкіндігі негізінде сөзтұлғаларды нормалау және қажетті грамматикалық ақпарат алумен қатар сөзтұлғаларды жинақтау (синтездеу) да іске асады (Марчук, 44).

Сөзтұлғалардың әріптік құрамының өзгеруін анықтайтын сөздерді өзара топтауды (топтарға бөлуді) негіз етіп алып, автоматтандырылған морфологиялық талдаудың алгоритмі құрастырылады. Мұндай топтау «морфологиялық топтау» деп аталып жүр (Белоногов, Новоселов). Сөзтұлғалардың жазылуындағы әріптік өзгеріске ұшырау флективті және агглютинативті тілдерде біркелкі емес.

Мәселен, орыс тіліндегі: «*сиджу — сидишь*» сөзтұлғаларының жазылуындағы сөз негізіндегі әріптер өзгеріске ұшыраса, сол сөздердің қазақша баламасы — «*отырмын — отырсың*» сөздерінің жазылуында қосымшадағы әріптер ғана өзгерген.

АМΠΑР атты машиналық аударма жүйесінде сөздердің синтаксистік қызметі мен септік және т.б. жалғаулардың негізінде сөздердің морфологиялық кластары екі түрге бөлініп қарастырылды:

- 1) негіздерінде өзгеріс болатын сөздер тобы,
- 2) сөздердің флективті тобы.

Соңғы топ белгілі сөздерге ғана тән қасиеттегі белгілер жүйесі арқылы немесе осындай қасиеттерді бойына сақтаған дерек сөздер арқылы сипатталады.

Енді морфологиялық талдаудың түрлеріне қысқаша тоқталайық. Олар төмендегідей:

- негіз сөздердің сөздігі арқылы морфологиялық талдау жүргізу;
- сөзтұлғалар сөздігі көмегімен морфологиялық талдауды жүзеге асыру;
- логикалық көбейту әдісін қолданып, морфологиялық талдау жасау;

— морфологиялық талдауды сөздіктердің көмегінсіз (сөздіксіз) арнайы кестелер арқылы жүзеге асыру.

Еуропа тілдерін зерттеуде көп тараған морфологиялық талдау түрі — негіз сөздердің сөздігі мен кейбір көмекші кестелер жүйесін пайдалану. Аталған сөздіктерде ішкі флексиясыз, жай және күрделі сөздердің негіздері мен олардың негіз тұлғалары (формалары) толық беріліп отырады. Сөздіктегі әрбір негіз сөзге морфологиялық кластардың екі түрін ажырататын шартты белгі (код) қойылады. Ал омонимдік негіздерге шартты белгілердің (кодтардың) тіркесімдік түрлері беріледі.

Морфологиялық талдаудың екінші түрі, яғни сөзтұлғалар сөздігі арқылы жүргізілетін талдау да зерттеушілер тәжірибесінде көп қолданыс тапқан, кең тараған талдау түрі болып саналады. Олай дейтініміз, морфологиялық талдаудың компьютерлік алгоритмінде морфемаларға бөлшектеу мен оларды сөздік бойынан іздестіру әрекеттері негіз сөздер сөздігіне қарағанда жеңіл жүзеге асады. Бірақ талдаудың бұл түрінің өзіне тән осал жерлері де жоқ емес. Мысалы, егер іздестіретін сөз сөзтұлға сөздігінде кездеспеген жағдайда, біз оның грамматикалық ақпараты жайлы да ештеңе біле алмаймыз. Сондықтан сөзтұлға сөздігі көмегімен талдау жүргізу жүйесінде аффикстер мен түбір сөздер тізімдері берілуі қажет. Осымен бірге бірнеше сөзтұлғаларды бір лексикалық бірлікке сәйкестендіруге қажетті, оларға тән қасиеттерге ие белгілер (атрибуттары) толық берілгені талап етіледі. Бұл айтылғандардан туындайтын қорытынды — морфологиялық талдаудың негіз сөз сөздіктері арқылы жүргізілгені ұтымды.

Морфологиялық талдауда «логикалық көбейту» әдісі айрықша орын алады. Бұл әдісте «сөздік функциясы» деген ұғымға ерекше көңіл бөлу керек. Әрбір «сөздік функциясына» сөзтұлға функциясы мен оған тән ақпарат сәйкес қойылады. Осының негізінде әр сөзтұлғаның өзіне ғана тән ақпараттық деректер арқылы функцияның мәнін арнайы кестемен беруге мүмкіндік туады. Бірақ аталған функцияның мәнін басқаша, төменде көрсетілгендей тиімді амалдар арқылы беруге де болады:

1) сөзтұлға әріптер тізбегі ретінде морфемдік сегмент-терге бөліктенеді;

2) сөзтұлға морфемдік сегменттер тізбегі ретінде басқа морфемдік элементтердің реттелмеген жиынымен ауыстырылады;

3) сөзтұлға морфемдік жиын ретінде белгілі-бір ақпарат сәйкес қойылады;

4) бұл ақпарат сөзтұлғаға қатысты деп саналатын қорытынды ақпарат түрінде қайта өзгертіліп беріледі (*Марчук, 46*).

Әрбір морфемаға құрылымында сондай морфемасы бар сөзтұлғалардың жиынтық ақпаратын да сәйкес қоюға болады. Логикалық түсінік бойынша, мұндай ақпараттар жиынтығы дизъюнкциямен, яғни сөзтұлғаның құрылымдық белгілерінің бір-біріне қарсы қойылуымен сәйкес келеді. Сөзтұлға жайлы ақпарат осы сөзтұлғаның табиғатына тән қиылысудан немесе логикалық конъюнкция деп аталатын морфемалық ақпараттардан тұратынын дәлелдеп шығуға болады.

Морфологиялық талдаудың логикалық көбейту әдісі, көбінде, флективті тілдерде қолданылады және негіз сөздер сөздігінің болуын қажет етеді.

Қарастырып отырған автоматтандырылған морфологиялық талдаудың соңғы түрі — сөздіксіз талдау немесе «тәуелсіз» талдау. Ол аффикстер кестелері көмегімен және арнайы түзілген грамматикалық мағынасыз сөздердің тізімі негізінде іске асады. Морфологиялық талдаудың бұл түрі зерттеу тәжірибесінде өте сирек қолданылады.

Енді «автоматтандырылған морфологиялық талдаудың қазіргі күйі қандай?» — деген сұрақтың жауабына қысқаша тоқталайық.

Қазіргі кезде кез келген мәтіннен компьютер арқылы сапалы ақпарат алу жүйесіне қойылатын талап заман сұранысына қарай өсті. Сондықтан компьютер көмегімен морфологиялық талдау жасау сызбасын (алгоритмін) жүзеге асыру мәселесінің негізі деп саналатын мынадай талаптар қойылады:

— кез келген тақырыптағы мәтіннің 98-99 пайызды қамтитын көптақырыпты (политематический) «аса күшті» сөздік морфологиялық талдау жүйесінің негізін құрауы тиіс;

— автоматты түрдегі талдау алгоритмі кез келген сөзтүрленістерін ескеретіндей мүмкіндікке ие болуы қажет. Осындай жағдай ескерілсе ғана сөзтұлғаны жан-жақты тану

мүмкіндігі артады және сан жағынан алғанда ондай мүмкіндіктің нәтижесі сөздікте қамтылған лексикалық бірліктен бірнеше есе артық болуы мүмкін;

— автоматтандырылған морфологиялық талдау жүйесінде «жаңа» сөздер, сөздікте қамтылған басқа сөздермен бірдей дәрежеде қарастырылуы қажет және оларды дұрыс танудың ықтималдығы 90-95 пайыздан кем болмауы тиіс;

— мәтінді компьютер арқылы өңдеу жылдамдығы тұтынушыны қанағаттандыратын болуы қажет және оған енгізілетін мәтін көлеміне шек қойылмауы керек;

— компьютер арқылы жүргізілетін морфологиялық талдау жүйесінде, қажеттікке қарай, өзгерістер жүргізу (сөздікті толықтыру, өңдеу және т.б.) мүмкіндігі алдын ала ескерілуі тиіс, яғни бұл аталған жүйені «оқытуға» болатындай мол мүмкіндік тууы керек (*Марчук*, 48-49).

Бүгінгі таңда, орыс тіліне және басқа да еуропа тілдеріне қатысты морфологиялық талдау жүйелеріндегі кейбір сөздіктер мен мәтіндердің көлемдері жайлы мынандай мәліметтерді айтуға болар еді:

— негіз сөздерден тұратын көптақырыптық сөздіктің көлемі 100 мың лексикалық бірліктен кем емес;

— сөздіктің көлемі 30 млн. сөзқолданыстан асатын көптақырыптық мәтін негізінде жасалғандықтан, оның кез келген тақырыптағы ғылыми-техникалық мәтінді қамту деңгейі өте жоғары;

— көлемі 3 млн. сөзқолданыстан тұратын мәтіннің негізінде түзілген сөзтұлғалар сөздігі 46 мың лексикалық бірлікті құрайды.

Сөз болып отырған екі сөздікте де сөздердің құрылымы мен синтаксистік сипаты жөнінде толық түрдегі грамматикалық ақпарат берілген (*Зеленков*).

Автоматтандырылған морфологиялық талдау сөздіктердің айрықша түрлерін өмірге келтірді. Оларда тілдік бірліктердің тұлғасына, туындау ережелеріне, сөзтудырғыш, сөзтүрлендіргіш топтарына қарай реттеу жақтары қарастырылды. Осындай типтегі сөздікке тілдік бірліктің соңғы жағынан бастап әліпби тәртібіне келтірілген «кері әліпбилі сөздік» деп аталатын сөздікті жатқызуға болады. Морфологиялық талдаудың талабына сай түзілген бұл сөздікте сөзтудырғыштық, сөзтүрлендір-

гіштік типтегі сөздер және соңғы қосымшалары бірдей болып келетін күрделі сөздер өзара жіктелген ретте көрініс табады. Кері әліпбилі сөздік бойынша сөздердің морфологиялық құрылымына қатысты көптеген мәселелерді анықтауға болады. Мәселен, қосымшалары бірдей болып келетін сөздердің топталып берілуі олардың грамматикалық сипатын және қосымша мен қандай сөзтүрлендіргіш типке жататындығы арасындағы қатынастарға қатысты деректерді анықтауға мүмкіндік туады.

Грамматикалық форманттардың синонимдік, омонимдік қатарларын жеңіл ажыратуға және олардың тіркесімдік, сандық мөлшерінің сипаттарын анықтауға мүмкіндіктер туады. Кері әліпбилі сөздік бойынша, аналитикалық тілдерге қарағанда флективті және агглютинативті тілдер жүйелерінен морфологиялық форманттарға байланысты ақпараттарды көптеп алуға болады.

Осы уақытқа дейінгі қазақстандық «Статистика тобының» қазақ тілін ЭЕМ-ның және компьютердің көмегімен лингвостатистикалық зерттеу барысындағы қол жеткізген көптеген ғылыми нәтижелерін, яғни әр мақсатпен түзілген жиілік сөздіктер мен олар арқылы алынған құнды деректер мен тұжырымдарды әрі қарайғы іс-тәжірибемізде дұрыс пайдалана білуіміз керек. Ең алдымен, сол ғылыми жетістігімізді үнді-еуропа тілдеріндегі «компьютерлік лингвистика» саласы бойынша орындалып жатқан ғылыми-зерттеу жұмыстарының деңгейіне дейін көтеру қажет.

Компьютердің араласуымен ғана жүзеге асатын морфологиялық талдаулар мен басқа да зерттеу түрлерін мемлекеттік дәрежедегі қазақ тілі аясына қолдана отырып, ана тіліміздің компьютерлік лингвистика саласын дамытуды және соның нәтижесінде қол жеткізген жетістігімізді күнделікті өмір қажеттігіне пайдалану ісі қазіргі заман талабы болып отыр.

Әдебиет

Баранов А.Н. Введение в прикладную лингвистику. Эдиториал УРСС. М., 2001.

Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. М.: Наука, 1979.

Василевский А.Л., Марчук Ю.Н. Вычислительная лингвистика. Учебное пособие для студентов отделения прикладной лингвистики. М.: МГПИИЯ им. М.Тореза, 1970.

Городецкий Б.Ю. Актуальные проблемы прикладной лингвистики / Новое в зарубежной лингвистике. Вып. XII., 1983.

Зеленков Ю.Г. Морфологический анализ в системах автоматической обработки научно-технической информации. Канд. дисс. М.: ВИНТИ, 1988. 145 с.

Лингвистические вопросы алгоритмической обработки сообщения. М.: Наука, 1983.

Марчук Ю.Н. Основы компьютерной лингвистики. Учебное пособие. М., 2000.

«Адам–компьютер–адам» жүйесіндегі табиғи тіл қызметінің ерекшеліктері

Қоғамның даму сатысы күрделене түскен сайын, ондағы ақпарат құралы қызметін атқаратын «тілдің» дәрежесі де айтарлықтай көтерілді. Қоғамның даму сипаты ақпарат ағымының үздіксіздігі мен сапалылығына өз әсерін тигізбей қоймады. Әсіресе, қысқа және ұзақ мерзімді ақпарат түрлері айтарлықтай өзгеріске ұшырады.

Осының әсерінен, біріншіден, тілдік құрылымның негізгі қаңқасы ретіндегі грамматиканың, түбір сөз бен өнімді сөзжасам моделі жүйесінің тұрақтылығын, екіншіден, өндірістік қажеттілік пен қоғам өмірінің қажеттігіне сай сол қаңқаны толықтыратын лексикалық бірліктер мен олардың тіркесімдерінің сан жағынан ұлғайуымен бірге, олардағы тұрақсыздық қасиеттерді де байқауға болады (*Андреев*).

Адам өміріне электрондық техниканың араласуы бұлшық ет күшіне қатысты көптеген жұмыстарды жеңілдетумен қатар, ми қызметінің де іс-әрекеттеріне араласа бастағаны бәрімізге мәлім. Осыған байланысты адам миының қызметіне бәсеке тудыратын неше түрлі «ақылды» автоматтар мен «ойлау» қабілеті бар электронды машиналар жайлы жайттар соңғы кезде көптеген басылымдарда жарық көруде. Кейбір авторлар, тіпті осыдан 200 жыл бұрын тоқу станогының тоқымашыны ығыстырғаны сияқты, электронды машиналар да адамзатты ми қызметінің өрісінен ысырып тастап, оған тек тұтынушының ғана рөлін қалдыруы мүмкін деп те күмән туғыза бастады. Әрине,

мұндай қисынсыз болжам адам баласының ми қызметінің мәнін дұрыс түсінбей, оны түгелімен бөлшектелетін (алгоритмделінетін) процеске жатқызып, компьютер арқылы оны қайта туындатуға болады-мыс деп ұйғарудан туындайды. Шындығында, адамның ойлау қабілеті негізінен дедуктивті-логикалық емес, индуктивті-ықтималдық болуынан, толығымен бөлшектеуге жатпайтын процесс деп ұйғару керек.

Бұл пайымдаулардан адамның ойлау қабілетін түгелімен формалдауға болмайтындығын және оны компьютерлік алгоритмдеу амалымен қайта туындатудың, яғни ойлау әрекетін қайтадан өзгеріссіз қалпына келтіру мүмкін еместігін ескеру қажет.

Соңғы кезде ақпараттар көлемінің өсуі жайлы жаңа көзқарас байқалады. Мәселен, ол – қысқа мерзімді ақпараттың көлемі арифметикалық прогрессия құрап өседі де, ал ұзақ мерзімді ақпарат – геометриялық прогрессия бойынша ұлғаяды деген көзқарастар. Сонымен қатар, ақпараттық процестерді компьютерлендіру (автоматтандыру) ғылым саласынан гөрі, өндіріс саласында көбірек және тезірек іске асады деп есептеу тенденциялары орын алып жүр.

Сол сияқты, табиғи тілге байланысты төмендегідей екі түрлі теориялық проблемаларды сөз етуге болады.

Олардың біріншісі – тілдің компьютер жадындағы қызметіне қатысты туындайтын сұрақ.

Табиғи тіл (ауызша не жазбаша) адам баласынан логикалық роботқа немесе компьютерге түгелдей ауысуы кезінде қандай сапалық өзгеріске ұшырауы мүмкін, әлде ондай «тіл» өзгеріссіз алғашқы қалпында қала ма?

Екіншісі – табиғи тілді зерттеу барысында формалдау процесі шекті ме? Басқаша айтқанда, «тілді» мазмұнға қатыссыз бөлек сипаттау мүмкін бе? Егер олай қарастыруға болмаса, ондағы тілдік бірліктердің берілуі мен мағыналарының арақатынасын анықтау мүмкін бе? – деген сұрақтар туындайды.

Бірінші сұрақтың жауабын «тіл» ұғымын әртүрлі тұрғыда қарастыруға байланысты түсіндіруге болады. Мысалы, «тілді» тек семиотика тұрғысынан анықтасақ, онда тіл – символдарға қатысты кодтар жүйесі деп саналады. Бұл түсінік бойынша адам да, логикалық робот та кибернетикалық жүйе деп ұйғарылады.

Осының негізінде «тілдің» адамнан роботқа немесе керісінше – роботтан адамға ауысу процесінде өзгеріс болмайды деп ұйғару керек.

Енді «тілді» мазмұнды лингвистикалық тұрғыдан қарастыратын болсақ, онда болмыста кездесетін «тілдің» екі түрі, яғни олардың жеке (дара) және ұлттық тіл түрлері болады деп ұғынған жөн. Тілдің кейбір қолданбалы мәселелерін шешуде семантикалық көзқарас ыңғайлы болғанымен, ол жеткіліксіз. Себебі табиғи тілдің екі жақтылық табиғаты оны таңбалар жүйесінің теориялық аясында толық баяндауға мүмкіншілік туғызбайды. Екі жақтылық табиғат, біріншіден, жекелік және ұлттық жүйе түрінде байқалса, екіншіден, коммуникативтік құрал ретінде көрініс табады.

Егер логикалық роботты (компьютерді) аса бір зор жазба кітапшасымен салыстыратын болсақ, оның ұқсастығымен қатар, айырмашылығын да аңғаруға болады. Ұқсастығы – екеуінде де табиғи тілдің жазба көрінісі болуында.

Компьютер мен адам арасындағы коммуникативтік жүйе арқылы мәтіндік ақпараттың адамнан компьютерге ауысу процесінде табиғи тіл оның көрінісі ретіндегі жазбамен алмасады да, ал керісінше, ақпарат компьютерден адам назарына ауысқанда, жазба мәтінінің компьютердегі көрінісі қайтадан таңбалардың психикалық кескін үйлесімінің бұрынғы қалпына келтіріледі, яғни ішкі тіл (внутренняя речь) қалпына ауысады.

Сонымен, кез келген мәтінмен компьютер арқылы жұмыс істеу – тілге психикалық әсер ету емес, ол мәтіннің формалды шартты белгілерінің оның өз таңбаларымен әрекет етуі деп түсіну қажет. Бұл әрекеттер, әрине, адамның ойлау қабілетіндегі процестерге ұқсай бермейді.

Табиғи тілдің құрылымын формалдаудың шегін анықтау қажеттігі тілдің берілу пішіні (план выражения) мен мазмұндық көрінісі (план содержания) арасындағы қатынастарын арнайы зерттеуді қажет етеді. Тілдік фактілердің формалдану байланыстарын зерттей келе, біз жазба тілдің берілу көрінісін баяндауда олардың ондай заңдылыққа бағынатынын байқаймыз. Ал тілдік элементтердің берілу көрінісі мен мазмұндық берілуінің арасындағы байланысты зерттеу барысында біз тікелей байқал-

майтын нысандарға келіп тірелеміз. Бұл жағдайда «тілдік белгі» мен «белгіні білдіруші» өзара психикалық пішін үйлесімінде ғана байланысты болмай, олар мазмұнды тұрғыдағы үнемі өзгерісте болатын экстралингвистикалық ұғымдармен ықтималдық байланыстарға да ұшырайды. Тілдің мазмұндық болмысының толық формалдануына кедергі жасайтын, негізінен алғанда, тілдегі семантикалық құбылыстар.

Осы айтылғандардан мынадай қорытынды жасауға болады.

Адам баласына тән сөйлеу не жазу тілі еш уақытта логикалық робот – компьютерге толық берілуі мүмкін емес. Компьютерді тілдің берілу көрінісін анализдеу мен синтездеу және мүмкіндігінше оларды формалдау жағдаяттарында пайдалануға болады. Ал мазмұнға тиісті зерттеулер белгілі бір семантикалық шектен аспайтын көлемдегі формалдануға тиісті жайттарда ғана іске асуы мүмкін (*Андреев*).

Компьютер көмегімен шешілетін жоғарыда сөз болған тілдік көріністерді формалдауға көнетін есептерді квазилингвистика саласына жатқызады. Мұндай есептер сан жағынан аз болғанымен, олардың қолданбалық шеңбері айтарлықтай мол болады. Квазилингвистикалық есептер адам өмірінің көптеген жақтарын қамтиды. Мысалы, олардың ойдағыдай шешілуі экономикаға байланысты жұмыстарды басқаруда, әскери қорғаныс істерінде және басқа да көптеген ғылыми зерттеу жұмыстарында маңызы зор деп саналады. Бүгінгі таңдағы осындай есептердің ең негізгісі – ақпараттық қор жинау. Бірақ бұл мәселені шешудің өзіндік қиындық жақтары бар. Ол, әсіресе, адам баласы мен компьютердің арақатынастарын (қызметін) дұрыс ажыратуға байланысты болып келеді. Мысалы, жиі қайталанатын және оңай формалданатын тілдік бірліктерге қатысты амалдарды: сұрыптау, топтастыру, санау, реттеу және т.б. жататын есептерді компьютерге жүктеген жөн болады. Ал стандарттауға жатпайтындар – семантикаға қатысты тілдік құбылыстар. Мысалы, мәтіннің мазмұнын ашуға қатысты – сөздердің көпмағыналығын айыру, омоним сөздердің мағынасын ажырату және сол сияқты кейбір формалдау процесіне көне бермейтін тілдік құбылыстар бар (*Андреев*).

Енді мәтіннің туындауы (порождение текста) жағдаятын және оны әртүрлі деңгейде қабылдаудың мүмкіншілігі жайлы қысқаша мәлімет берейік.

Мәтін өзара қарым-қатынаста болатын тілдік жүйелердің «норма», «узус» деп аталатын ережелері мен шектеулеріне және тілдің өзіне байланыссыз пайда болатын сырт жағдайларына байланысты туындайды. Мұндағы «сырт жағдай» деп отырғанымыз, мәтінде баяндалатын оқиға мен шектеулерге (узусқа) қатысты контекстік жағдаяттар (*Пиотровский*).

Қабылдаушы адамның өзіне бағытталған хабардың мән-мағынасын анықтауы, оның тілдік жүйені меңгеру дәрежесіне және тілдің ережелері мен шектеулерін («норма» мен «узус») айыра білуіне, әрі сол хабарламадағы сөз болатын жағдайлармен қаншалықты таныстығына да көп байланысты.

Қабылдаушы адам мәтіннің мән-мағынасын жан-жақты ажырата білуі, оның тілдік жүйені және оның «норма», «узус» деп аталатын қасиеттерін толық игеруімен қатар, осы айтылымда пайда болған жағдаяттың прагматикалық құндылығын ескеруіне де қатысты.

Енді хабарламаны лингвистикалық тұрғыдан қабылдау деңгейлерін сөз етер болсақ, онда мынадай сатыларды ажырата білуіміз керек:

1) грамматикалық деңгей (бұл деңгейде тек сөйлемнің грамматикалық схемасы ғана қабылданады);

2) сөздік деңгей (мұнда сөзтұлғалардың лексикалық мағынасы ғана қабылданады).

3) сөздік деңгейдегі жоғарғы саты — ол қабылдаудың фразеологиялық деңгейі. Бұл қабылдауда сөз тіркестерінің мағынасы ашылады, ал тіркес арасындағы грамматикалық байланыстар ескерілмейді.

4) келесі саты — жоғарыда аталған үш түрлі қабылдауды жалпылайтын саты — мәтін бойындағы лексика-грамматикалық деңгейді айқындау. Бұл сатыда да мәтін толық түрде лингвистикалық қабылдаудан өтпейді. Себебі қабылдаушы адам аталған деңгейлерде тек қана сөздер мен сөз тіркестерінің мағынасын ашып, мәтіннің грамматикалық құрылымын анықтағанымен, оның тереңде жатқан семантика-синтаксистік (әрі стильдік) байланыстарын аңғармай қалуы мүмкін. Мұндай

байланыстарды анықтау үшін қабылдаушы адам мәтіндегі семантика-синтаксистік деңгейге көтерілуі қажет. Бұл деңгей, әрине, қабылдаушының тілді толық меңгеруін талап етеді.

Қысқаша айтқанда, болашақ автомат-роботтың тілдік білім дәрежесі әлеуметтік деңгейде болуы тиіс. Бірақ мұндай дәрежедегі электронды автоматты құрастыру ісі, әзірше, көптеген күтпеген арақайшылықтарға ұшырауда. Осындай қиыншылықтардың негізгісі — адам баласының сөйлеу тіліндегі дискреттік (бөліктенулік) және үздіксіздік тәрізді қасиеттері болып табылады (*Пиотровский*).

Мәтінге қатысты инженерлік-лингвистикалық жұмыстар, яғни мәтінді толық қабылдау және оны «туындату» (порождение) тек «адам—компьютер—адам» (A1—K—A2) коммуникативтік жүйесінде ғана іске асады. Бұл жердегі адам, яғни «A1» — хабаршы, «A2» — хабарды қабылдаушы рөлінде, ал компьютер, яғни «K» — хабаршының хабарламасын қабылдаушы адамға жедел және қолайлы түрде жеткізетін аралық компонент.

«Адам—компьютер—адам» — хабардың берілу жүйесінің схемасы төмендегідей:



Суреттегі A1—K—A2 хабардың берілу жүйесін екіге: (A1—K) және (K—A2) бөліп қарастыруға да болады. Бірінші процесте (A1—K) мәтіндік хабарлама адамнан компьютерге беріледі, ал екіншісі — (K—A2), керісінше, компьютерден адам назарына қарай аударылады. Осы екі түрлі процестің әрбіреуін төмендегідей төрт блокқа бөліп қарастыруға болады (*Пиотровский*).

Бірінші процесс A1—K :

1.1) хабаршының (A1) назарында (ойлау процесінде) хабарлама мәтінінің «туындауы»;

1.2) «туындалған» мәтін белгілерінің (кодтары) импульстер тізбегі ретінде компьютерге берілуі;

1.3) қабылданған импульстер жиынтығының компьютерлік семиозис негізінде мәтінді айқындау және оның мағынасын (мазмұнын) формалдау;

1.4) компьютер жадындағы айқындалған мәтінді мақсатқа сай өңдеу (өзгерту), мысалы, басқа тілге аудару, мәтіннен

аннотация, реферат алу, индекстеу және нәтижесін табиғи тіл түрінде қабылдаушыға ұсыну.

Екінші процесс **К–А2** төмендегідей төрт топтамадан тұрады:

2.1) компьютер жадындағы мақсатқа сай өңделген ақпаратты тұтынушыға ұсыну (бұл топтама жоғарыдағы көрсетілген 1.4-ке ұқсас);

2.2) компьютерлік нәтижені қабылдаушы адамның (А2) қабылдауы;

2.3) компьютерлік хабарламаның семиозисі;

2.4) қабылдаушы адамның (А2) хабарламадан алған әсері.

Хабардың берілу жүйесі қызметінің ең маңыздылары деп, компьютерлік семиозисті (1.3), мәтінді айқындауды және қабылдаушының компьютер ұсынған нәтижеге байланысты түсінігін айтады. Осы аталған жағдайларда да семиозистің жетістігі табиғи тілдің компьютерлік тілге, яғни жасанды математикалық тілге ауысу нәтижесінің дәрежесіне тығыз байланысты.

Инженерлік тіл білімінің келесі бір маңызды проблемасы – лингвистикалық алгоритмдер мен олардың программаларын құру мәселесі. Мұны басқа сөзбен – “компьютерді оқыту” деп те айтуға болады. Табиғи тіл мен жасанды тілдің қатынастық жақтарын дұрыс түсінбейінше, бұл мәселені шешу оңайға түспейді.

Табиғи тіл де, жасанды математикалық (компьютерлік) тіл де семиотикалық жүйеге жататындықтан, олардың негізгі қасиеттеріне тоқталайық.

Тіл жүйесін тілдік бірліктер және олардың арасындағы қатынастар деп түсінетініміз мәлім. Ал табиғи тілдің бірліктері дегеніміз: дыбыс пен фонема, әріп, морфа мен морфема, сөз бен сөзтұлға және олардың құрылымдық түрлері, сөз тіркесі, сөйлем және олардың қаңқалық схемасы. Енді жасанды компьютерлік тілдің бірліктері – әріп, цифрлар мен символдар және олардың шартты белгілері (кодтары), компьютерлік түбір сөз, сөздер және сөйлемдер. Бұл жердегі компьютерлік сөз не түбір сөз деп отырғанымыз, белгілі бір түрде шектелген әріп тіркестері, ал компьютерлік сөз тіркес және сөйлемдер дегендеріміз, алдын

ала келісім бойынша шектелген компьютерлік сөздер тізбегі болып табылады.

Табиғи тілдің мынадай үш түрлі функцияларын атауға болады (*Баранов*):

а) семантикалық және сигматикалық ақпаратты таратуға мүмкіндік туғызатын лингвистикалық символды қалыптастыратын тілдің номинативті функциясы;

б) тілдік белгілерді өзара байланыста ұстайтын және мәтіндік ақпараттың синтактикалық қасиетін сипаттайтын тілдің предикативті функциясы;

в) хабардың берілу кезінде тілдің номинативті және предикативті функцияларын хабарлаушы (А1) не қабылдаушы (А2) жеке адамның (жеке ұйымның) қалауымен ұштастыра алатын прагматикалық функция.

Жасанды математика тілінде және оның барлық тармақтарында, яғни компьютерлік тілдің берілуінің өзінде-ақ предикативті функция біте қайнасып жатады.

Жасанды компьютерлік тілдің номинативті функциясы белгінің (знактың) мағынасымен (десигнатымен) шектеледі. Ал бұл жасанды математикалық тілде прагматикалық функция атымен жоқ деуге болады (*Пиотровский*).

Сонымен, хабардың берілу жүйесі (А1–К–А2) жұмысының нәтижелілігі, біріншіден, мазмұнды хабардың көлемдік мөлшеріне және екіншіден, көпшілік қабылдаушының дайындық дәрежесіне де көп байланысты демекпіз.

Қысқаша айтқанда, инженерлік тіл білімінің жалпы тіл теориясына қатысты атқаратын кейбір қызметтері осындай.

Әдебиет

Андреев Н.Д. Язык в обществе, использующем кибернетические машины//Вопросы социальной лингвистики. Л.: Наука, 1969. С. 25-40.

Баранов А.Н. Введение в прикладную лингвистику. Эдиториал УРСС. М., 2001. 360 с.

Пиотровский Р.Г. Текст, машина, человек. Л.: Наука, 1975. 327 с.

Жасанды интеллект жүйесін құрастыруға қатысты мәселелер жайында

Қазіргі түсініктегі электронды-есептеу машина (ЭЕМ) деген сөз тіркесі ең алғаш 1946 жылдан бастап (ЭНИАК, АҚШ) қолдана бастады және бұл есептеу техникасының қарқындап даму кезінің бастамасы еді. Міне, осы үрдіс қазіргі кезде де өз жалғасын табуда.

Бірінші есептеуіш машиналар көлемді арифметикалық есеп-қисаптарды орындауға арналып, ол машиналар программалар арқылы әрекет ететін ғажайып үлкен калькуляторлар тәріздес болды. Мысалы, осындай ЭНИАК машинасы баллистикалық кестелерді есептеу үшін жобаланған болатын. Бірақ өткен ғасырдың 50-жылдарының басында-ақ ЭЕМ-ды пайдалану мұндай шағын аямен ғана шектелмеу керектігі анық бола бастады, яғни есептеуіш машиналардың ақпарат түрінің саннан да басқаша түрлері бойынша әрекет ету (өңдеу) мүмкіндігі бары анықталды. Дәлірек айтсақ, практикалық тұрғыдан қарастырғанда, кез келген ақпарат (мәтін, сурет, дыбыстық сигнал және т.б.) ЭЕМ көмегімен өңделетін сандар арқылы кодталуына (шартты сандық белгімен ауыстырылуына) мүмкіндігі бары айқындалды. Бірақ бұл жайт практикалық әрекеттер үшін айтарлықтай маңызды емес еді. Оның себебі қазіргі кездегі есептеу техникасы мен оны пайдалану әдістерінің даму сипатына қарай ЭЕМ үшін қажетті деген программаны жазып, мысалы, қайсыбір мәтінді өңдеуге қатысты қойылған міндеттерді шешуге болады. Ал бұл мәтіннің қалайша кодталғаны

(шартты белгіленуі) туралы мәліметті білуді тіпті де қажет етпейді.

Есептеу техникасының мүмкіндігін (яғни ЭЕМ арқылы сандық емес ақпаратты өңдеу мүмкіндігін) тез арада ұғынуға математикалық логиканың қарқынды түрде даму барысы себепші болды. Мысалы, ең алғашқы есептеу машиналарының пайда болуынан бұрынырақ кезде алгоритмдер теориясының (математикалық логиканың тарауы) Тьюринг машинасы, рекурсивтік функциялар, Марковтың нормалды алгоритмдері және т.б. нәтижелерге қол жеткен болатын. Сонымен, XX ғ. 50-жылдарынан бастап қазіргі кезге дейін электронды есептеу машиналары сандық емес ақпараттың да әртүрлі мәселелерін өңдеуде кең қолданыс тауып жүр.

Өткен ғасырдың 50-жылдарының өзінде-ақ ғалымдарды мынадай сұрақтар мазалайтын: машинаны адамша ойлатуға бола ма немесе, басқаша айтқанда, жасанды интеллекті құрастыру мүмкін бе?

Шынында да, тірі адам ағзасындай ойлау қабілеті бар жасанды құралды ойлап табу мәселесі адамның тарихи даму кезеңдерінде бірнеше рет қойылғаны мәлім. Мәселен, алхимиктердің гомункулусты алу әрекеттерінен бастап, ежелгі және ортағасырлық аңыздармен жалғаса отырып, Карел Чапек пен Айзек Азимовтың роботтарына дейінгі мәселелерді мысалға келтіруге болады. Бірақ тек ең алғашқы есептеу машиналарының пайда бола бастауынан-ақ бұл мәселе, абстрактылы философиялық сызба мен көркем шығармалар деңгейінде емес, белгілі дәрежеде болашақтың ісі бола тұра, ақиқаттықты ескеретін деңгейде нақтылық түсінік ала бастады. Бұл мәселе ең бірінші рет ағылшынның тамаша математигі А.Тьюрингтің 1950 жылы жарық көрген мақаласында нақты тұжырымдалып, жан-жақты сипатта қарастырылды. Мақала «Машина ойлай ала ма?» деп аталған болатын.

Жасанды интеллекті құрастырып шығуға бола ма? — деген сұраққа жауап алу, бір қарағанда, жеңіл болғанымен, ол аса күрделі мәселе еді. Оның күрделілігі мынада: біз әлі де болса адам баласының ойлау жағдаятын дұрыс зерттей алмадық және ойлаудың жалпыға ортақ сипаттағы теориясы әлі де жоқ деуге болады. Оны айтпағанның өзінде, біз әлі «ойлау» деген не? —

деп қойылатын сұраққа да жауап бере алмаймыз. Осы уақытқа дейін “ойлау” процесінің жасанды интеллект мәселелерін шешуге қатынасы бар философтарды, психологтарды, инженерлерді, тілшілерді және т.б. мамандарды бірдей дәрежеде қанағаттандыра алатын анықтама жоқ.

Мәселен, “ойлау” қабілеті бар жүйе, яғни жасанды интеллект жүйесі құрастырылды деп есептелік. Сонда оның «ойлай алатынын» немесе «ойлай алмайтынын» қалайша тексеруге болады? Интеллектуалдықты қалай өлшейміз? Осы мәселелерге байланысты А.Тьюрингтің жоғарыда аталынған тұңғыш мақаласында мынадай өлшем ұсынылып және ол ғылыми әдебиеттерде «Тьюринг тесті» деген атпен аталып кетті. Қысқаша түрде оны былайша тұжырымдауға болады: егер адам белгілі уақыт аралығында электронды машинамен қарым-қатынас жасап отырып, өзінің адаммен емес, машинамен «тілдесіп» отырғанын байқамай қалғанын сезінетін болса, тек сондай жағдайда ғана машинаның ойлай алатындығын мойындауға болады. Мұндай жағдайда адамның машинамен «тілдесуі» тікелей жүзеге аспауы мүмкін. Мысалы, ақпаратты жолдауда адам қайсыбір құралды (телетайпты) пайдаланады. Әрине, мұндай шектеу қою барынша табиғи екені белгілі. Себебі, ол интеллектуалдық өлшемі ішінен «сыртқы көрініс», «дауыс тембірі» және т.б. деп аталатын өлшемдерді (параметрлерді) жоққа шығаруы мүмкін:

Тьюринг тестісінің өзіндік субъективтілігіне қарамай, ол интеллектуалдық өлшем ретінде зор құндылыққа ие болды. Оның ең үлкен құндылығы – адамның ойлау әрекетімен салыстыру әдісін негіз етіліп алынуы. Әзірше ойлаудың жалпы теориясының болмауынан «бірдеңенің» ойлау қабілеті бар немесе жоқ деп мойындау үшін оны тек адаммен ғана салыстыра аламыз. Себебі, ол мұндай қасиетке әлем бойынша иегер жалғыз ғана тірі жан болып саналады.

Сонымен бірге аталған Тьюринг тесті барынша әмбебеп қасиетке ие. Сұрақ және жауап әдісі практикалық тұрғыда адам баласының интеллектуалдық әрекетінің кез келген саласын қамтуға мүмкіндік туғызады. Тьюринг тестісіне қосымша шектеулер енгізе отырып, біз интеллектуалдық өлшемінің кейбіреуін ғана туындата аламыз, яғни ол жасанды интеллекттің

кейбір жеке міндеттерін ғана шеше алады деген сөз. Мысалы, біз машина мен адам арасындағы «сұхбаттың» белгілі пәндік сала аясынан ауытқымауын талап етуімізге болады. Сол сияқты арадағы қарым-қатынас тілінің лексикасы мен синтаксисіне де және т.б. шектеулер қойып, тілдесуді жүзеге асыруға болар еді.

Жасанды интеллектің қандай жеке міндеттерді шеше алатындығы жайлы Тьюринг тестісінің ешбір көмегінің болмауынан, зерттеу мәселелерінің аясы интуиция бойынша ғана анықталады. Мұндай міндеттер топтамасын әр зерттеуші өзінше анықтаса да, оның «тұрақты өзегінің» болатындығы белгілі. Олар — тәжірибе жүзінде, көпшілік мамандардың жасанды интеллект шешетін міндеттер санына қосатын мәселелері: есептеу әрекетін орындау, шешім қабылдау, бейнелерді тану.

Есептеу әрекетін жүзеге асыру. Бұл термин көптеген мәселелердің басын біріктіреді: жоғары алгебраның теоремаларын дәлелдеу мен орта мектептің физика оқулықтарындағы есептердің шешімін табудан бастап, күнделікті қажеттілікті өтейтін мәселелерге дейін. Мысалы, тіпті “қолымыздағы бар азық-түліктен қандай тағам жасауға болады” деген сұраққа жауап алуға дейін. Әрине, бұл жерде аталған есептердің нақты шешімін табу ғана емес, осы тәріздес есептердің жауабын іздеудің жалпы механизмін табу, яғни есеп шығару барысының моделін табу жайлы сөз болып отыр.

Әзірше, негізінен, әртүрлі математикалық есептерді модельдеуге қатысты мәселелер зерттелуде. Себебі, мұндай математикалық есептер формалдау әдісіне тезірек көнеді. Сонымен бірге аталған мәселенің жалпылама шешімін табу үшін адамның пайымдау, қисынды қорытынды жасау мүмкіндіктерін зерттеп, түсініп алу қажет. Тек математикалық есептерді модельдеу үшін ғана формалды логика пәнінің әдістерін қолдануға болар еді. Әрине, егер мұндай модельдеуге математикалық логика әдістерінің “күші” жетсе. Себебі адам баласы қолданатын логикалық әдістер формалды логикадан тіпті де басқаша болып келеді. Бұл айтылғаннан адамдар формалды логиканы еш пайдаланбайды деген қорытынды шықпайды, бірақ олар нақты түрде қисынды өзгерістер жүргізу мақсатында тағы басқа да, әзірше белгісіз, «бірдеңе»

қолданатыны анық байқалады. Мүмкін, тек формалды логика ғана үстемдік ететін әлем өте-мөте көңілсіз болатын шығар. Сонымен, егер математикалық логиканы барынша жақсы зерттелген ғылыми пән деп есептесек, онда «адам логикасы» тіпті де зерттелмеген сала деуге болады.

Шешім қабылдау мәселесі, негізінен, стратегиялық ойындар түрлерінде қарастырылады. Мысалы, шахмат, дойбы және т.б. ойындарды модельдеуге қатысты жұмыстар тек жұртшылықтың көңілін көтеру үшін ғана жүргізілмейді. Дұрысы, керісінше. Шахмат және басқа да ойындарды жүзеге асыратын коммерциялық бағдарламалардың шын мәніндегі мақсаты бөлек. Олай дейтініміз, «ойын ережелері» тек тақта бетіндегі шахмат фигураларын бір орыннан екінші орынға ауыстырып отыру үшін ғана емес екен. Оның түпкі мақсаты, мысалы, шикізат пен құрал-саймандарды өндіріс орындарына орналастыру, әскери сахнадағы іс-әрекеттер үшін полктер мен армияны жылжыту, финанстық жүйе бойынша қаржының орнын ауыстыру және т.б. сол сияқты мәселелердің шешімін табуға сай ұқсастықтар деуге болады.

Әрбір нақты жеке жағдайға байланысты ең жақсы нәтижеге әкелетін ережелерді тандау негізінде **шешім қабылдаудың** ортақ сипаттағы әдіс-тәсілі анықталады. Осының нәтижесінде тек тиімді автоматтанған эксперттік жүйелерді құруға ғана қол жетіп қоймайды, сонымен бірге ойлау әрекетінің жалпы теориясына да белгілі дәрежеде үлес қосылады деуге болады.

Бейнелерді тану. Адамға өзін қоршап тұрған материалдық әлемді үнемі топтастырып отыру әрекеті тән екені мәлім. Адам әртүрлі нысандардың шексіз жиынын топ-топқа бөлшектеп қарастыруға дағдыланған. Мысалы, екі алманың өзара ұқсастығының аздығына қарамай, адам баласы, оны бір сөзбен — *алма* деп атайды. Ал *алмұрттың алмаға* өте ұқсас болуына қарамай, оны басқа сөзбен — *алмұрт* деп атайды.

Сол сияқты, демалыстан оралған жақсы таныс адамның өзгеріп — күнге күйіп, сақал-мұрты өсіп, қара көзілдірік киіп алғанына қарамай, біз оны басқа адаммен шатастырып алмай танимыз.

Әртүрлі адамдардың бір әріптің өзін ұқсастырмай әртүрлі жазғандарына қарамай, біз олардың бір әріп екендігін тани

аламыз және бір ғана сөзді бірнеше адамның дыбыстау әрекетінің ерекшеліктеріне қарамай, біз оны да бір сөз деп ұғынамыз.

Екі түрлі нысанды теңдестіру үшін адам баласы қандай көзге шалына бермейтін белгілерге сүйенеді? Яғни, мысалы, екі заттың да алма екендігін анықтау үшін немесе кешегі көрген адам мен бүгінгі көріп тұрған адамның бір адам екендігін айқындау және т.б. үшін адам баласы қандай белгі-мүмкіндікті пайдаланады?

Ойлау барысында адам баласы нысандар мен олардың топтарының қайсыбір эталондық бейнесін құрастырады деп жорамалдауға болады. Тануға тиісті нысанды адам өз санасында орын алған бейнемен салыстыру нәтижесінде ол нысанды бейне топтарының біреуіне жатқызады. Бұл жердегі «бейне» сөзін қайсыбір психологиялық немесе нейрофизиологиялық термин деп ұғынудан гөрі, интуитивті түрде, адамның сол нысанға байланысты қайсыбір түсінігі немесе нысан топтары жайлы оған таныс бірыңғай сипаттамалар деп ұғынғанды жөн демекпіз. Әрине, нақты зерттеулерде «бейне» ұғымы дәлірек айқындалуы мүмкін.

«Бейне тану» мәселесінің шешімін табудың негізгісі – ол ең күрделі мынадай екі сұраққа жауап іздестіруді қажет етеді: эталондық бейне қалайша құрастырылады (яғни адам қалайша әлемді бөліктейді және топтайды) және эталонға қарағанда нысан бейнесінің қандай айырым белгілері мәнді, ал қандайлары мәнсіз?

Адам өзін қоршаған ақиқаттық жайлы ақпаратты әртүрлі арналар бойынша алуға мүмкіндігі бар – көру, есту, сезіну және т.б. Сондықтан адамдарда бейненің әртүрлі типтерінің қалыптасуы мүмкін болғандықтан, бейнені тану ісі бірнеше мәселелерге ажыратылады: көру бейнелерін тану, акустикалық бейнелерді тану және т.б.

Тьюринг тестісінің жасанды интеллект құрастыруға қажетті мәселелерді шешуге байланысты көмегінің болмайтындығы жайлы жоғарыда да айтқанбыз. Солай бола тұра, жасанды интеллект мәселелерінің Тьюринг тестінде аталған ең бір маңызды бағыты – адам мен машина (компьютер) арасындағы қарым-қатынасты (тілдесуді) табиғи тілде жүргізуді қамтамасыз

ету. Бұл мәселенің маңыздылығы тек жасанды интеллектті құрастыру ісіне біршама жақындату мен ойлаудың жалпы теориясына белгілі үлес қосу ғана емес, сонымен бірге толық түрде шешілмесе де, оның көптеген практикалық қолданыста пайдаға асатындығында. Мысалы, ол диалогтік ақпарат-ізвестіру жүйесінде, басқару ісінің автоматты жүйесінде және басқа да арнайы тілдерді үйренуді және арнаулы сөздіктерді қолдануды қажет ететін жүйелерді құрастыруда аса маңызды.

Сонымен, қолданбалы тіл білімінде маңызды деп саналатын жасанды интеллекттің көптеген мәселелерінен адамның сөйлеу әрекетін модельдеу проблемасы, сөйлеуді талдау мен жинақтау мәселелері жеке қарастыруды қажет етеді.

Адам мен машинаның табиғи тілде қарым-қатынас жасау барысын былайша елестетуге де болар еді: машина мәтінді өз жадына енгізіп, қабылдап, өз тарапынан белгілі түрде тиісті әсер (реакция) етеді. Мұндай әсер құрастырылатын жүйенің міндеті мен мүмкіндіктеріне қарай әртүрлі болуы да мүмкін. Мәселен, ол синтезделінген машиналық мәтін түріндегі сұраққа жауап беру әрекетінің әсері болуы немесе жүйенің ішкі жағдайының өзгерісі (мәтіндегі ақпаратты жауапсыз қабылдау) мен қайсыбір жай ғана әрекет (роботтың өз иесінің бұйрығын орындауы) түрінің әсері болуы да мүмкін. Осындай мүмкін болатын әсерлердің ішінен біз тек сұрақ-жауап жүйесі мәселелерімен ғана шектелеміз.

Мәтінді қабылдау кезінде машинаға мәтін «ұғынықты әсер» етуі үшін ең алдымен машина оны «түсінуі» қажет болады. Осыған байланысты мынадай сауал туындайды: мәтінді «түсіну» дегеніміз не және машинаның мәтінді дұрыс түсінгенін қалайша анықтауға болады? Бұл аталған сұрақтар аса күрделі сұрақтарға жататыны белгілі. «Ойлау» (мышление) терминіне анықтама берудің қиындығы сияқты, «түсіну» терминіне де қатысты анықтама жоқ деуге болады. «Ойлау» мен «түсіну» терминдерінің бір-бірінен алшақ еместігі жайлы ғана жорамал жасауға болады.

Зерттеушілердің «түсіну» критерийін туындатуы Тьюринг тестісіне ұқсастық сипатта қарастырудың нәтижесі болуы да мүмкін. Мәселен, егер машинаның мәтінге қатысты әрекеті барабар және дұрыс сипатта болса ғана машина мәтінді түсіне

алады. Бұл критерийдің субъективтілік сипаты (Тьюринг тесті тәріздес) айқын: машинаның мәтінге әсер ету әрекетінің дұрыс-бұрыстығын кім анықтайды? Сонымен бірге аталған критерийде “түсіну” ұғымында “дұрыс түсіну” жайлы да сөз болып отыр. Яғни бұған “бұрыс түсіну” жағы да қатысады деген сөз. Қалай болғанда да жасанды интеллекттің практикалық жүйесін құру жағдайында бұл критерийдің қажеттігін терістеуге болмайтындығы белгілі.

Мәтінді дұрыс түсінуге болатындығы жайлы аталған критерий де (Тьюринг тестісі сияқты) көмекші құрал бола алмайды. Бұл мәселенің шешімін табуға байланысты әр зерттеушінің өз ұстанымы болатындығы белгілі. Сонда да бұл бағытқа жалпылама түрде былайша сипаттама беруге болады: жасанды интеллект жүйесі мәтінді қайсыбір ішкі көрініс құрылымына қайта өзгертуі қажет, яғни мәтін мағынасы құрылымының көрінісіне келтіруге байланысты өзгертулер жүргізіледі. Осы құрылымды негіз етіп ала отырып, аталған жүйе өзінің «интеллектуалдық қызметін» атқара алады. Мәселен, қойылған сұраққа қатысты қажетті деген жауапты ұйымдастыру, қайсыбір әрекетті жүзеге асыру және т.б.

Ішкі көрініс құрылымының көпшілік мақұлдаған ешбір де ережелері мен нұсқаулары жоқ. Бар болғаны тек бір ғана барынша күнгірт сипаттағы талап – ішкі көрініс құрылымының мәтін мән-мағынасын дұрыс ашу мен әртүрлі «интеллектуалды» әрекеттерді жүзеге асыру мүмкіндіктеріне ие болу. Сонымен бірге әртүрлі мән-мағынаға әртүрлі құрылымның сәйкес келуі.

Кейбір зерттеушілер жасанды интеллекті жүйесінің мәтінді түсіну барысы «адамның түсіну» деңгейіндей болуын талап етеді. Ондай ғалымдардың түсінігінше, басқаша жағдайда табысқа жету мүмкін емес. Әрине, бұл талап та өзінше қисынды деуге болады. Бірақ бүгінгі таңда мәтінді дұрыс түсіне алатындығын мақтаныш ете алатын тірі жан иесінің бірі және бірегейі – ол тек адам. Сондықтан адам баласында ғана орын алатын мұндай жағдаят (процесс) осы сала бойынша оң нәтижеге әкелетін жалғыз-ақ мүмкіндік деуге болады. Бірақ жасанды интеллекті жүйесіне адам мүмкіндігіне сай келетіндей талап қою ақиқат шындыққа жараса бермейді. Әрбір зерттеуші адам өз түсіну мүмкіндігін сенімді түрде негіздей алмайды.

«Адамның басы» зерттеушілерге қайсыбір «қара жәшік» тәріздес. Ал адам басындағы түсінік алу кезінде болатын жағдаяттарды барынша дәл айқындауға нейролингвистика ғылымының жетістіктері әлі де болса жеткіліксіз. Сондықтан мамандар өз зерттеулерінде аталған мәселенің «адам басында қалайша шешім табылатынын» жүзеге асыруға ниеттенгендерімен, олар сөйлеу қызметін функционалдық тұрғыда модельдеу қажеттігі жайлы аса сақтықпен айтуда.

Сонымен, жасанды интеллект жүйесі мәтінді “түсіну” үшін ол ең алдымен мәтінді ішкі көрініс құрылымына ауыстыруы қажет. Сонда ғана мәтіннің мән-мағынасы қажетті түрде бейнеленеді. Мұндай ауыстыру жұмысын жүзеге асыру үшін жасанды интеллекті жүйесі қайсыбір ақпаратқа сүйенуі керек болады. Әдетте жасанды интеллектімен айналысатын мамандар бұл ақпаратты *білім* деп атап жүр. Көпшіліктің мақұлдауы бойынша *білім* екіге бөлініп қарастырылады – *тіл туралы білім* және *әлем туралы білім*. Білімнің соңғы түрін *энциклопедиялық білім* деп те атайды.

Аталған *білім* түрлері аралығына дәл шекаралық сызық жүргізу барынша күрделі, тіпті мүмкін де емес деуге болады. Әрбір зерттеуші бұл мәселені өзінше шешуге тырысуда. Көбінде тіл туралы *білімге* морфологиялық, синтаксистік, семантикалық ақпараттарды жатқызады. Бұл аталғандар мәтіннің құрылымы мен сөздер аралығындағы негізгі байланыстар арқылы мәтін құрылымын анықтауға мүмкіндік туғызатын мәліметтер. Ал мәтінді түсінуге қажетті барлық басқа ақпараттар *әлем туралы білім* түріне жатады.

Табиғи тілдегі мәтінді түсінуге арналған алғашқы эксперименттік жүйелерді құрастырушылар тек *тіл туралы білімдермен* ғана қанағаттануға талаптанған болатын, бірақ көп ұзамай-ақ мұның жеткіліксіз екендігі, яғни *қоршаған орта жайлы білімнің* де қажеттігі айқындалды.

Әртүрлі адамдардың *әлем туралы білімі* айтарлықтай дәрежеде айырым табуы мүмкін. Сол сияқты мұндай жағдай әртүрлі жасанды интеллект жүйелеріне де қатысы бар. Ал бұл айтылғандар мәтінді түсінуге өз әсерін тигізбей қоймайды. Сондықтан да әр адамның бір ғана мәтінді әртүрлі (өзінше) түсінуі орын алады. Бұл айтылғандарға мәтін үзінділерінен

көптеп мысалдар келтіруге де болады. Міне, мұның өзі жоғарыда сөз болған критерийдің субъективтілігінің тағы бір дәлелі деуге болады.

Тіпті ең қарапайым деген мәтінді дұрыс түсіну үшін айтарлықтай көп көлемдегі *әлем жайлы білімді* пайдалануымыз керек болады. Егерде *тіл туралы білімді* азды-көпті дәрежеде толық сипаттауға мүмкіндік туатынына қарамай, *әлем туралы білім* қорының шексіздік сипаты оны толық қамтуға мүмкіндік туғызбайды. Солай бола тұра, адам баласының әлемді тану мүмкіндігі де күннен-күнге артып келетінін ескеру қажет. Осыған байланысты түсіну жүйесін құруда *әлем туралы білімге* қатысты кейбір принциптік мәселелер туындайды. Мәселен, барлық білім қорын қамту мүмкін болмағандықтан, жұмыс барысында олардың ең қажеттілерін алдын ала болжау да аса күрделі. Құрастырылып жатқан жүйелердің негізгі бөлігі аса шағын пәндік салаға бағышталған, сондықтан ондай жүйелер тек шектеулі тақырыпты (мазмұнды) мәтінді ғана түсіне алады. Бірақ осындай шектеу жасаудың өзінде де алынған нәтижелер шындықтан әлі де болса алшақ.

Аталған мәселеге қатысты келесі бір зерттеу бағыты — оқыту жүйелерін (обучающие системы) құруға байланысты. Бұл бағыттың болашағының зор екені белгілі. Себебі, оның түсінік алу жүйесі арқылы мәтін ішінен жаңа ақпаратты тауып, бұрыннан жинақталып жатқан *әлем туралы білім қорына* қоса алатындай мүмкіндігі бар. Егер мәтінді түсінуге қажетті ақпарат жеткіліксіз болып жатса, оның күңгірт (түсініксіз) жерлерін анықтау үшін ондай жүйе тұтынушы адамға қосымша сұрақтар туындатып, білім қорын жинақтауға мүмкіндігі бар. Бірдей көлемдегі білім қоры бар жүйелер алғашқы кезеңде әртүрлі нәтижелерге ие болуы мүмкін.

Бірақ әлем туралы білім қорын жинақтау жолдарына байланыссыз (тәуелсіз), мәтінді түсіну кезіндегі негізгі міндет — мәтіннің ішкі көрінісінің құрылымын, яғни мәтіннің мән-мағынасының құрылымын құрастыру. Әдетте мұндай құрылым мәтінді автоматты түрде өңдеу кезіндегі әрекеттерге ұқсас болып келеді. Мәселен, машиналық аударма, автоматты рефераттау және т.б. орын алатын морфологиялық, синтаксистік және семантикалық талдау әрекеттеріне аса ұқсас болып келеді.

Құрастырылған құрылымды жүйенің қарсы әрекетін туындату үшін де қолдануға болады. Оған жататындар: жауапты тұжырымдау, әрекет түрін таңдау және т.б. Әрине, аталған кезеңге қатысты зерттеулерді жүзеге асыру тек тілшілердің үлесіне ғана тиісті емес, сонымен бірге басқа да көптеген пәндер өкілдерінің тікелей араласуын қажет етеді. Мысалы, психологтардың – әрекет үлгісін анықтауда, логика мамандарының – қисынды қорытынды жасауда және т.б. жасанды интеллект саласындағы барлық зерттеу бағыттары практикалық тұрғыда әртүрлі мамандықтардың мүдделеріне жақын келеді, сондықтан олар әртүрлі пәндік мамандарды қатыстыра отырып, кешенді түрде жүзеге асуы қажет.

Жоғарыда сөз болған мәселелердің бәрін бірдей қанағаттандыратын ортақ шешім әлі де болса табылмай келеді. Әрине, жеке бір мәселеге қатысты аса тар шеңбердегі мәселелердің шешімдерін табуға арналған талпыныстар көптеп кездесетіні белгілі. Мәселен, егер адам мен компьютер аралығындағы тілдесу табиғи тілде жүргізілетін болса, онда ол шектеулі пәндік сала бойынша, ал егер нақты шешім қабылдау үшін модельдеу әрекеті қажет болса, онда ол нақты мәселенің шекаралығында (шахмат ойынындағы жүрісті таңдау) жүзеге асады және т.б.

Сонымен, бүгінгі таңдағы ойлау әрекетінің жалпы теориясы жаңадан ғана дамып келеді деп қорытынды жасауға болады.

Әдебиет

Моделирование языковой деятельности в интеллектуальных системах / Под ред. А.Е. Кибрика и А.С. Нариньяни. М., 1987.

Панков И.П. Искусственный интеллект//Прикладное языкознание: Учебник. отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. С. 91-99.

Попов Э.В. Общение с ЭВМ на естественном языке. М., 1982.

Фитиалов С.Я. Понимание естественного языка в системах искусственного интеллекта//Лингвистические проблемы функционального моделирования речевой деятельности. Вып. V. Л., 1982. С. 3-9

Хант Э. Искусственный интеллект. М., 1978.

Оқу лексикографиясының ғылыми пән ретіндегі ерекшелігі

Егемен еліміздің орта және жоғары оқу орындарында салалық пәндерді мемлекеттік тілде оқыту аса күрделі және кезек күттірмейтін көптеген мәселелерді шешуді талап етіп отыр. Осыған байланысты басты мәселенің бірі – сөздіктер, оқулықтар мен оқу құралдарын өмірге келтіруде ғылыми педагогикаға негізделген қазақ тілінің жалпы лексикалық қорын және салалық пәндердің терминологиясын құрастыру.

Қазақ тіліндегі әр салалық пәндерге қатысты терминдерді қалыптастыруда ғалым-лексикографтарды орыс және басқа да туыстас түркі халықтарының озық тәжірибелеріне сүйенетін күрделі жұмыстар күтіп тұр деуге болады. Атап айтар болсақ, халық ауыз әдебиетінің бай мұрасы мен қазақ халқының тұрмыс-тіршілігіне, салтына тән көптеген ежелгі атаулар мен терминдерді оқырманның жас ерекшелігіне, білім деңгейіне қарай сөздіктер мен оқу құралдарында орынды пайдалана білу ісі жүзеге асар еді.

Қай ғылым саласын алмасақ та, оның өзіне тән метатілі болатыны мәлім. Метатіл ұғымына тек терминдер ғана емес, жалпы ғылыми лексика да енеді. Ал мұндай метатілдің қалыптасуына сөздіктердің де әсері болатындықтан, көпшілік ғалымдар тіл білімінің лексикография саласын теориялық және

практикалық негіздемесі бар ғылыми пән ретінде қарастыруды жөн санайды.

Ахмет Байтұрсынұлы мен Құдайберген Жұбанов бастап дамытқан қазақ тіліне қатысты терминдер белгілі дәрежеде қалыптасқанымен, басқа ғылыми пәндердің қазақша терминдік жүйесі әлі де жинақтауды, зерделеуді қажет ететіні белгілі. Сондықтан да оқулықтарды, сөздіктерді пайдалана отырып, олардың түсініктемесін, түсіндірілуін айқындау қажет болып отыр.

Қазіргі кезеңде қазақ тілінің сөздік қоры жаңарып, тіл білімінің арнасы кеңеюіне байланысты оқулықтар мен сөздіктер де уақытқа сай өзгеріске ұшырауда. Жарық көріп жатқан сөздіктерде соңғы жылдары пайда болып жатқан қазақ және орыс тілдеріндегі жаңа сөздер мен байырғы сөздердің түсіндірмелері мен қолданыстары енгізілуде. Сол сияқты ондай сөздіктерде қазақ әдеби тілінің нормаларына сәйкес және күнделікті қарым-қатынаста, оқыту жүйесінде жиі қолданылатын сөздер де орын алуда. Айтылуы қиындық тудыратын қазақ немесе шет тілінің сөздері мен сөз тіркестерінің транскрипциясы беріліп, реестр сөздерінің стильдік ерекшеліктері қалыптасқан ескертпелер мен шартты белгілер арқылы көрсетіп беру жұмысы да атқарылуда. Өртүрлі мақсатты ұстаным ететін сөздіктерде сөздің қолдану сипатын жан-жақты қамту үшін кеңінен таралған қазақтың мақал-мәтелдері, сөздердің негізгі және ауыспалы мағыналарын айқындайтын сөз тіркестері келтіріліп, омонимдер мен көпмағыналы сөздер қатарларына ерекше көңіл бөлу ойластырылуда.

Әрине, қазақ тілінің лексикография саласының дамуына тіл білімінің төрінен орын алған мынадай іргелі еңбектер негіз болғаны мәлім: «Қазақ тілінің түсіндірме сөздігі» (1-10 том, А., 1974–86), І.Кеңесбаевтың «Қазақ тілінің фразеологиялық сөздігі» (А., 1977), Н.Т.Сауранбаевтың жетекшілігімен жарыққа шыққан «Казахско-русский словарь» (М., 1954), Р.Сыздықтың жетекшілігімен жарық көрген «Жаңа атаулар» (А., 1999), А.Байтұрсынұлы атындағы Тіл білімі институты лексикографтарының құрастыруымен және жалпы редакциясын Т.Жанұзақовтың басқаруымен шыққан «Қазақ тілінің сөздігі»

(А., 1999), «Қазақша-орысша сөздік» (А., 2001), «Орысша-қазақша сөздік» (А., 2005) және т.б.

Қазақстан Республикасы Үкіметінің жанындағы Мемлекеттік терминологиялық комиссиясының әр жылдары ресми бекіткен түрлі сала терминдері жинақталып, дүркін-дүркін жеке сөздік түрінде шығарып тұру да дәстүрге айналды. Осындай лексикографиялық жұмыстардың нәтижесі ретінде 1976, 1979, 1990 жылдары тілші ғалым А.Әбдірахмановтың құрастыруымен шыққан «Атаулар мен терминдер» сөздігі, 1996 жылы жарық көрген Б.Қалиев пен Ш.Құрманбайұлының «Орысша-қазақша және қазақша-орысша терминдер сөздігі», Тілдерді дамыту департаментінің 1998 жылы Т.Алпысбаев пен А.Әбдірахмановтың және 1999 жылы жарық көрген құжаттар жинағының басын қосып, 2002 жылы басылып шыққан «Терминологиялық жинақ», сондай-ақ салалық терминологиялық орысша-қазақша, қазақша-орысша сөздіктердің 31 томы «Рауан» баспасынан жарық көрді (2000 ж.). Осы аталғандардың қатары-на 1998 жылы «Ғылым» баспасынан шыққан Б.А.Жұбановтың 25677 сөзді қамтитын көптілді (ағылшын, неміс, орыс және қазақ тілдері бойынша) «Химиялық сөздікті», А.Салқынбай мен Е.Абақанның авторлығымен «Сөздік-словарь» баспасынан 1998 жылы шыққан «Лингвистикалық түсіндірме сөздікті», Т.Момынов пен А.Рақышевтың 1999 жылы «Атамұра» баспасы-нан жарық көрген орысша-қазақша «Медициналық сөздікті» жатқызуға болады. Сол сияқты М.Қазбеков пен Р.Шойбековтың «Русско-казахский словарь» атты сөздіктің (А., 1996), 2002 жылы «Аруна» баспасы шығарған 30000 сөзді қамтитын мектеп оқушылары мен жоғары оқу орындары студенттеріне арналған «Сөздіктің» (10000-нан аса сөз, құрастырғандар: Б.Исмағұлова, Э.Ережепова, Г.Әбдіжапарова) және 2006 жылы Астанада жарық көрген З.Ө.Жантекеева мен О.Нақысбековтың «Қазақша-орысша және орысша-қазақша оқушыларға, студенттерге арналған сөздіктің» (14000 сөз) де қазақ және орыс тілдерін оқып-үйрену кезінде аса құнды лексикографиялық еңбектер екені белгілі.

Осы бағыттағы аталған және басқа да аталмай қалған көптеген басылымдардың қазақ тілін оқыту процесінде, әсіресе жоғары білімді, білікті мамандар дайындауда игі көмегін

тигізері сөзсіз. Қазіргі жағдайда оқу процесін қамтамасыз ету үшін арнайы маман-педагогтар мен ғалым-лексикографтардың араласуымен қазақ тілінің оқулықтарын, оқу құралдарын, көптілді аудармалы, түсіндірмелі және энциклопедиялық сөздіктерін сапалы құрастыру заман талабы деуге болады. Өкінішке қарай, мектеп оқушылары мен студенттерге арналған қолымызда бар бірқатар сөздіктерде, оқулықтар мен оқу құралдарында әлі де болса теориялық тұрғыда кеткен кейбір олқылықтар та жоқ емес.

Сондықтан оқырман назарына ұсынып отырған еңбегімізде осындай кемшіліктерге болашақта жол бермеу тұрғысынан оқу лексикографиясына қатысты кейбір теориялық және практикалық мәселелерді сөз етуді жөн көрдік.

Оқу лексикографиясы жалпы лексикографияның жас саласы болып саналатыны белгілі. Солай бола тұра, оның да өзіне тән жақсы зерттелген тарихы мен дәстүрі бар. Ресейде ең алғашқы оқу лексикографиясының орталығы тек 1966 жылы ғана құрылып, Мәскеудегі А.С.Пушкин атындағы Орыс тілі институтында сектор ретінде қызмет етеді. Қазіргі кезде аталған сектор орыс тілінің біртүлдік және екітүлдік (орысша-шеттүлдік) сөздіктерін құрастырудың теориялық және практикалық жұмыстарын кең көлемде атқаруда.

Оқу лексикографиясының ғылыми пән ретіндегі өзіне тән ерекшелігі, оның жалпы лексикография, лексикология, семасиология психо- және әлеуметтік лингвистика және т.б. салалардың тілді үйрету әдістемесіне қатысты мәселелерімен тығыз байланысты болуында. Сондықтан оқу лексикографиясы дәстүрлі жалпы лексикографияның мәселелерімен қатар, мынандай көптеген әдістемелік факторларды ескеруді қажет етеді: оқыту кезеңдеріне қатысты сөйлеу әрекетінің түрі, тілді үйренушілер үшін тілдің ана тілі (немесе аралық-тіл) екендігі және қажетті тілдік ортаның болу-болмауы, жас ерекшелігі мен жалпы мәдени деңгейге қатысты болуы және т.б. мәселелер.

Осы аталған лингвистикалық және әдістемелік факторлар тек бірдей (және кешенді) дәрежеде ескерілсе ғана оқу лексикографиясын пайдалануға байланысты, ана тілін не шет тілдерін үйретудегі оқу сөздіктеріне қатысты дұрыс әдістемелік нұсқау беруге болады. Лексикографияның жалпы теориясы принцип-

терін орынды пайдалану негізінде тілді үйрету жолындағы ескірген әдістерге жол бергізбей, ішкітілдік және тіларалық баламалардың шегін анықтауға, лексика-семантикалық тіркесім және т.б. жайттарды айқындауға мүмкіндік туындайды.

Әдістемелік факторларды ескеру негізінде мына жайттар анықталады (Марусенко, 1996, 309):

1) сөздіктің адресаты (мектеп оқушысы, студент және т.б.);

2) сөздіктің типі (біртүлдік немесе аударма);

3) сөзтізбе көлемі (минимум сөздік, орта немесе толық);

4) сөздік мақаланың құрылымы, ондағы лексикалық материалдың берілу әдіс-тәсілі (эквиваленция, дефиниция, безендірілу, лексика-семантикалық топтағы басқа сөздерге қарама-қарсылығы), синонимия, анонимия, паронимия, ассоциативтік, тіркесімдік, грамматикалық, орфоэпиялық және стильдік ерекшеліктері және т.б.

5) полиграфиялық орындалуы.

Лексикографияның жалпы теориясын дұрыс ескеру тілдік ақиқаттықты шынайы көрсете білуге мүмкіндік жасайды. Ал методикалық талаптар оқу сөздігіндегі ең маңызды деген жайттарды, мысалы, сөздіктің тілді үйренуге бағытталуын, белгілі оқыту кезеңі мен оқыту формасын қолдануын, яғни оның қызметін анықтайды.

Оқу лексикографиясының тәуелсіз лингводидактикалық пән болуының мәні, оның тіл білімі мен әдістеменің өзара тығыз байланыстылығы мен өзара қатыстылығында жатыр.

Шынында да, егер қазіргі түсінікте, оқу лексикографиясы тіл білімінің ең бір жас саласы бола тұра, практикалық тұрғыдан алғанда оны бұрыннан келе жатқан сала деуге де болады. Бұлай тұжырымдау үшін кез келген сөздікті, ең алдымен, дидактикалық шығарма деп түсінген жөн болар еді. Лексикограф, яғни сөздік авторы оқулықтың авторы тәріздес, қоғам (ұжымдық тіл иесі) мен жеке адам аралығындағы делдал ретінде қызмет етеді. Сондықтан кез келген сөздік типінің қызметі – жеке адамның санасындағы білім мен ұжым санасындағы білім аралығындағы айырмашылықты жоюға талпыныс жасайды деуге де болады.

Қолданбалы тіл білімінің ірі маманы В.Г. Гак сөздіктердің қызметін мынадай негізгі төрт топқа бөліп қарастырады:

1. Тіларалық қатынасты ұйымдастыруға байланысты. Мұндай қызмет көбінде екітілді сөздіктер мен тілдескіштер арқылы іске асады.

2. Тілді оқыту (үйрету) үшін және мағынасы қиын сөздерге түсініктеме беру арқылы.

Сөздіктің мұндай қызметінде лексикографиялық материалдар ретінде біртүлді немесе екітілді сөздіктер, глоссариялар, тілдескіштер қолданылуы мүмкін. Оқыту процесі мынадай мәселелерге қатысты болады:

а) сөйлеу тілінен айырымы бар көне мәдениет тіліне қатысты тілдерді меңгеруде. Мысалы, ол ежелгі дәуірге қатысты тіл немесе басқа да тілдерге – латын, ежелгі грек тілі, Еуропадағы ортағасырлық ескі словян тілі, Шығыс елдеріндегі классикалық араб тілі және т.б. тілдер болуы мүмкін;

ә) шет тілдеріне;

б) ана тілге.

3. Ана тілін сипаттау мен нормалауға байланысты. Мұндай қызметтер, ең алдымен, түсіндірме сөздіктер арқылы және орфографиялық, қиын сөздер сөздігі және т.б. арнаулы біртүлдік сөздіктер арқылы іске асады.

4. Тілдерді зерттеу ісіне байланысты. Мұның негізінде біртүлді және екітілді филологиялық сөздіктердің әртүрлі типтері құрастырылады (Гак, 1977).

Лексикографияның аталған қызметтерін тарихи тұрғыда қарастырсақ, оның даму сатыларын былай жіктеуге болар еді:

1. Сөздікке дейінгі кезең (Еуропада – XV ғ. дейінгі кезең). Бұл кезеңде сөздік нәтижелері глоссариялар, вокабулалар (хрестоматияға не оқулық тарауына қатысты берілетін қысқаша сөздік), ономастикондар түрінде көрініс тапты. Олардың негізгі мақсаты мәдениет тілінің түсінуге қиын сөздерінің мағынасын ашу.

2. Сөздік түзудің алғашқы кезеңі (Еуропада – XV–XVII ғасырлар). Бұл кезеңде лексикондар (белгілі қызмет аясына ғана тән сөз қоры) мен тіларалық қарым-қатынас және мәдениет тілін зерттеу мен шет тілін оқыту үшін қызмет ететін екітілдік сөздіктер өмірге келді.

3. Лексикографияның дамыған кезеңі (Еуропада – XVIII ғ. бастап). Бұл кезең сөздіктердің әртүрлі типтерінің, әсіресе,

түсіндірме сөздіктердің шыға бастауымен ерекшеленеді (Марусенко, 1996, 310). Құрастырылған сөздіктер лексикография саласының жоғарыда аталып өткен төрт түрлі қызметін іске асыруды мақсат етеді.

Сонымен, оқу лексикографиясы сөздік құрастырудың алғашқы даму кезеңінің өзінде-ақ пайда болды деуге болады. Себебі, бастапқы кезендердегі сөздіктердің аттарының өзі-ақ, олардың дидактикалық міндеттерін нұсқап отырды. Мәселен, бірінші латынша-ағылшынша сөздік «*Promptulorium parvulorum*» деп аталған, яғни «Балаларға арналған сөз қойма». Ерте-ректегі французша-латынша сөздіктер – «*Commentarius puero-rum*», яғни «Жастарға арналған түсіндірме сөздік» деп және т.б. аталған. Бері келе мұндай сөздіктердің оқуға бағыттағыш сипаты екінші орынға ауысып, біріншіге – түсіндірме, тарихи, этимологиялық сөздіктер типтеріне ауыса бастайды.

Лексиканы оқыту мақсатының нысаны етіп алуға арналған әдебиеттерді талдай келе, ғалым В.В.Морковкин мынандай негізгі бес аспектіні бөліп қарастырады (*Актуальн. проблемы ...*, 1977, 31-35 бб.):

1. Оқу сөздіктері. Мұндай сөздіктерге қатысты мәселелер, әсіресе, сөзтізбе құрылымын талдау мен оқу сөздіктерінің типтері жайлы мәселелер толық зерттелді деуге болады. Бірақ бұл зерттеулердің негізгі мақсаты әр типті сөздіктерде орын алатын ұқсастық пен айырмашылықтардың аналитикалық мәселелері, лексиканың семантикасын анықтаудың тәсілдері және тағы басқалар.

2. Лексикалық минимум сөздіктері. Қазіргі кезде тілдің сөздік қорын минимизациялау теориялық жағынан айтарлықтай зерттелді деуге болады. Соның нәтижесінде оқу формасына қарай көптеген лексикалық минимум сөздіктері құрастырылды. Бірақ оқу процесінде лексикалық минимумның барлық артықшылығын орынды пайдалану үшін ол жан-жақты сипатта жүзеге асуы және оқу барысының барлық кезендерін түгел қамтуы қажет. Әрине, бұл аталған мәселелердің әлі де толық шешімі табылмай келеді.

3. Лингвостатистикалық зерттеулер. Мұндай зерттеулердің нәтижелері, әдетте, жиілік сөздіктер мен үлестірімді жиілік

сөздіктер (қолдану жиілігі әр стиль бойынша үлестіріліп берілген) түрлерінде көрініс табады.

4. Лексикографиялық типтегі лексикаға қатысты оқу құралдары. Өртүрлі тілдерге арналған мұндай оқу құралдары көптеп шығарылды деуге болады. Бірақ олар, теориялық тұрғыда, әлі де болса зерттеу нысаны бола алмай отыр.

5. Оқулықтар. Мұндай оқулықтарда лексиканың семантикалық жақтары әр тақырыпқа қатысты пәндік сөздіктерде, жинақтық (сводный) сөздіктерде және арнайы лексикалық жаттығуларда қарастырылады. Қандайда болсын оқулықтар үшін лексикалық аспект өте-мөте маңызды деп саналады. Бірақ мұндай оқулықтарда лексиканы семантикалау тәжірибесінің теориялық негіздемесі орын ала бермейді.

Жоғарыда көрсетілгендей, оқу сөздігі, ең алдымен, қажетті әдістемелік бағытталуымен, нақты оқу кезеңдеріне қатыстылығымен, оқудың мақсаты мен міндетіне және т.б. қарай сипатталады. Оқу сөздіктері тілдік материалды тандау ісінде арнайы мақсат қоюмен ерекшеленеді. Яғни болашақ лексикографиялық талдау үшін қажет болатын тірек бірліктердің негізгі тізімін түзу ісімен ерекшеленеді. Сөздіктің сөзтізбесі әрі шағын, әрі жеткілікті сипатта болуы қажет. Бұл мақсат тек сөздің негізгі мағынасындағы квантитативті параметрлерін, грамматикалық тұрпатын және синтаксистік құрылымын, тақырыпты ашатын ең маңызды сөздер мен сөз тіркестерді тандауды қатаң түрде есепке алу арқылы ғана жүзеге асады. Осымен бірге сөздіктің негізгі тақырыбы мен сөйлеу жағдаятын қамтуды тексеру қажеттігі де туындайды. Өкінішке қарай, тәжірибе жүзінде мұндай тексеру жұмысы көбінде орын ала бермейді.

Оқу сөздігінің сөзтізбе көлемі оның белгілі бір типіне қарай ғана анықталады. Мысалы, қысқаша түсіндірме сөздіктегі лексика осы тілдік ұжымның әр мүшесі үшін қалыпқа айналған сипатта болуы қажет. Аталған мәселені шешу үшін мәдениеттің және ғылым мен өндірістік білімнің орта жиынтығы қамтылған ең кіші сөйлеу қызметінің моделі құрастырылуы керек. Мұндай түсіндірме сөздік-анықтағыш 20000–25000 сөздік мақаладан тұратыны белгілі. Бірақ мұндай сөздік пассивті немесе жеке индивидтердің активті сөздік қорының шекаралығында тұрады да, оқу сөздігінің активті типіне жатады.

Кез келген оқу сөздігі қатаң мақсаттылықпен жүргізілген таңдаудың нәтижесі, яғни кең түсініктегі лексикалық минимум бола тұра, ол ішкі диалектикалық қарама-қарсылыққа толы. Бір жағынан, ондай сөздік нақты бағыттағы мәтіндерді және оқыту кезеңіне қатысты кейбір қиындықтарды да қамтуы қажет. Ал екіншіден, ондай сөздіктерде кез келген мәтінді түсінуге мүмкіндік туғызатын нақты материалдар да орын алуы керек. Бір ғана принципті (жиіліктік, тақырыптық, ситуативтік және т.б.) ұстаным ететін сөздік типі қойылған талапты орындай алмайды.

Сөздіктің әр деңгейіне тән сөздер бірнеше топ құрайды. Оларға жататындар (Марусенко, 1996, 313):

а) тілдің грамматикалық қаңқасын құрайтын және кез келген мәтінде кең қолданыс табатын – құрылымдық сөздер;

ә) көптеген мәтіндерде қолданыс табатын және сол себепті нақты тақырыпқа, сөйленіс жағдаятына байланыссыз келетін – жиілігі жоғары негіз сөздер;

б) белгілі тақырыппен немесе жағдаятпен тығыз байланысты болатын – тақырыптық, жағдаяттық сөздер. Бұл категорияға қатысты сөздер жоғары жиілікке ие емес, шектеулі сөйлеу мәнмәтіндерінде қолданылады. Сөздікке тек әртүрлі категориядағы (құрылымдық, тақырыптық, жағдаяттық) бірліктерді енгізу ғана кең шеңбердегі мәтіндерді түсінуге мүмкіндік туғызады.

Оқу сөздігінің құрылымдық бөліктерінің өзгешелігі. «Үлкен» (ірі) сөздіктерге қарағанда оқу сөздіктерінің құрылымы негізгі екі ерекшелікпен сипатталады: лексикалық бірліктерге түсініктеме берумен (семантикалау) және ол бірліктердің парадигматикалық, синтагматикалық қасиеттері мен қатынастарын анықтайтын әдіс-тәсілдерімен.

Лексикалық бірліктерді семантикалау түсіндірме оқу сөздігінің маңызды құрылымды бөлігі болып табылады. Әрине мұндай компонент белгілі дәрежеде басқа да сөздіктер типіне де қатысты. Аударма оқу сөздіктерінде семантикалау, әдетте, аударма баламалары арқылы, ал түсіндірме сөздіктерде мұндай семантикалау – анықтамалар арқылы жүзеге асады. Оқу сөздіктерінің маңызды құрылымдық бөлігі саналатын анықтама:

шағындық, көрнекілік және қарапайым сөзге саятындай талаптарға сай келетіндей болуы керек.

Шетел азаматтарына арналған түсіндірме сөздіктерде қайсыбір ұғымға түсініктеме беретін анықтамаларында таныс сөздердің шағын көлемді бөлігін пайдаланған жөн болады. Бұл сөздер «минимум сөздік» құрайды және осылардың негізінде сөздіктің басқа да лексикалық бірліктері анықталады. Мысалы, кең танымал М.Уэстің (*West, 1965*) сөздігінде 24000 сөздің мағынасы көлемі 1500 сөздерге тең лексикалық топ мағыналары арқылы анықталады. Сол сияқты француз тілінің негізгі лексикалық тобын қамтитын (*Gougenheim, 1959*) анықтағышта 55 сөз арқылы және қарапайым сөздіктің 1374 сөздері бойынша басқа сөздердің мағыналық анықтамалары берілген.

Оқу сөздіктерін құрастыру тәжірибесінен байқалатын жайт, ол «минимум сөздіктер» негізінде түзілген сөздіктер, көбінде, жасандылық пен нақтылықтың жеткіліксіздік сипатымен өзгешеленеді. Бірақ мұндай олқылықтар семантикалау әдісінің кемшілігінен емес, құрастырушылардың жіберген қателігінен деуге болады. Оқу сөздіктері үшін аса маңызды мәселе сөздердің синтагматикалық және парадигматикалық байланысын бейнелеп көрсете білу. Үйренуге тиісті тілдің тек лексикасын ғана меңгеру сөйлеу тіліндегі оның дұрыс қолданылуына кепіл бола алмайтындығы белгілі. Ана тілі мен шет тілінің түйісу әсерінен үйренушінің тілінде бұрын-соңды қолданылмайтын кейбір сөз тіркестерін туындатады. Ірі филологиялық сөздіктерде сөзқолданыстардың ерекшеліктері жайлы, яғни сөздердің тіркесімдігі жайлы мәліметтер бола бермейді. Сондықтан тәжірибе жүзінде, осындай мәліметтер бере алатын, сөздердің ең ықтимал нормативті байланысын бейнелейтін оқу сөздіктері қажет-ақ (*Регинина и др., 1980*).

«Парадигма» дегеніміз бір-біріне қарама-қарсы қойылған, бірақ ортақ белгісі бар немесе ұқсас ассоциация тудырушы кез келген лингвистикалық бірліктер тобы екені белгілі. Міне, осындай парадигматикалық қатынастарды бейнелейтін сөздіктерде тақырыптық, синонимдік, антонимдік топтар көрініс табады.

Оқу сөздіктерінің типологиясы. Типологиялық тұрғыда оқу сөздіктері бірнеше көзқараста қарастырылады:

- 1) сөздіктің қызметі тұрғысынан;
- 2) сөздіктің экстралингвистикалық параметрі бойынша;
- 3) сөздіктің ішкі лингвистикалық параметрі бойынша.

Оқу сөздіктері қызметіне қарай, жалпылама және мамандандырылған сөздіктер деп те ажыратылады. Бірінші сөздік үйренушілердің барлық категорияларына арналады да, ал екіншісі олардың тек белгілі категорияларына ғана арналады. Мысалы, мектеп оқушыларына, филолог-студенттерге, филолог емес студенттерге, мұғалімдерге және т.б.

Сөздіктің экстралингвистикалық параметрлері әрбір сөздіктің өзіне тән ерекшелігін анықтайды. Мысалы, олардың энциклопедиялық сөздікке жақындығына қарай немесе оның кейбір элементтерінің ғана энциклопедиялық сөздікке енгендігі бойынша. Осы тұрғыда мынадай бірліктерді бөліп қарастыруға болады: материалдық мәдениеттің заттық сипаты, рухани мәдениеттің концептері, ел тану компоненті және үйренуге (зерттеуге) қажетті тілдегі логикалық категориялардың бейнелену сипаты. Яғни «әлем бейнесінің» әртүрлі жекелік, топтық, ұлттық санасының концептуалды бейнеленуі. Экстралингвистикалық параметрлер жалқы есімдердің, мақал-мәтелдердің, қанатты сөздердің (сөйлемшелер) және т.б. сөздіктері үшін аса маңызды рөл атқарады.

Ішкі лингвистикалық параметрлер негізіндегі сөздіктер типологиясы тілдің сөздік құрамының типологиясын бейнелеуі қажет. Кез келген тілдің сөздік құрамы жүздеп-мыңдап саналады. Сондықтан қажетті сөзді тез іздеп табу үшін лексиканың жүйелік сипатын ескеретін стратегияны туындатып алған жөн болады. Соның негізінде қажетті деген сөзді барлық сөздік құрамының ішінен іздемей, тек оның кішігірім бөлігінің ішінен, мысалы, семантикалық өріс ішінен немесе лексика-семантикалық топ ішінен және т.б. іздеу мүмкіндігі туындайды. Оқу лексикографиясының теориясы, ең ұтымды жағдайда, тіл жүйесінің лексикалық аспекті бойынша, аспектілік сөздіктер жүйесін құруды қалайды. Сол арқылы іздеудің табиғи стратегиясы модельденіп, қажетті сөзді іздеп-табу мезгілі қысқарады.

Бұл мәселе сөздіктің көлеміне, оларды ұйымдастыруға, сөздіктегі әрбір лексикалық бірлікті сақтайтын ақпарат көлеміне, әрқилы таңбалар жүйесіне, сілтемелер жүйесін құру

жұмыстарына қатысы бар. Яғни бұл мәселе кешенді (жан-жақты) сипатқа ие деуге болады.

Қазіргі кезде халықаралық қатынас нығайып, көптеген дербес ұлттық мемлекеттердің пайда болуы және бұрынғы отарлықта қалған халықтардың мәдениетінің көтерілу сипаты, қайтадан лексикографияның тіларалық қарым-қатынасқа және тілдерді оқытуға байланысты қызметтерін бірінші орынға ауыстыруына жағдай жасайды. Әлемдік өндірісті интеграциялау кезеңінде әртүрлі түсініктерді бірізге түсіру ісі де аса қажет болуда. Сондықтан егемен Қазақстанның халықаралық деңгейге шығуына және бірқатар жетекші елдермен қазіргі таңдағы ғылым мен технологияның әртүрлі салаларында ғылыми және өндірістік ақпараттармен алмасудың едәуір ұлғаюына байланысты неғұрлым кең ауқымды қамтитын көптілді сөздіктер шығару қажеттілігі де туындауда. Жақын болашақтағы лексикографиялық зерттеулердің негізінде оқу сөздіктерінің әр ұлттық тілдерге, соның ішінде қазақ тіліне қатысты да практикалық маңызы зор жүйесі құрылуы мүмкін.

Сонымен, оқу лексикографиясының ғылыми пән ретіндегі ерекшелігі жайында жоғарыда сөз болған мәселелер және де осы бағытқа қатысты басқа да теориялық жайттар ескерілетін болса, қазақ тілінің оқу лексикографиясына байланысты көптеген әрекеттерді компьютер көмегімен толық автоматтандырып, олардың электронды түрлерін көптеп шығару жүзеге асар еді.

Әдебиет

Актуальные проблемы учебной лексикографии. М., 1977.

Гак В.Г. О некоторых закономерностях развития лексикографии (учебная и общая лексикография в историческом аспекте) // Актуальные проблемы учебной лексикографии. М., 1977.

Жантекеева З.Ө., Нақысбеков О. Қазақша-орысша және орысша-қазақша оқушыларға, студенттерге арналған сөздік. Астана, 2006. 208 б.

Жұбанов Б.А. Химиялық сөздік ағылшын, неміс, орыс және қазақ тілдерінде. Алматы: Ғылым, 1988. 200 б.

Исмағұлова Б., Ережсенова Э., Әбдіжапарова Г. Оқушыларға, студенттерге арналған қазақша-орысша, орысша-қазақша сөздік. Алматы: Аруна, 2002. 416 б.

Марусенко М.А. Учебная лексикография // Прикладное языкознание: Учебник / Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. 528 с.

Момынов Т., Рақышев А. Медициналық сөздік (орысша-қазақша). Алматы: Атамұра, 1999. 312 б.

Регина К.Б., Тюрина Г.П., Широкова Л.И. Устойчивые словосочетания русского языка. М., 1980.

Салқынбай А., Абақан Е. Лингвистикалық түсіндірме сөздік. Алматы: Сөздік-Словарь, 1998. 304 б.

Gougenheim G. Dictionnaire fondamental de la langue française. Paris, 1959.

West M. An International Reader's Dictionary. London, 1965.

Ғылыми-техникалық лексикография және арнаулы сөздіктер түрлері

Қазақ тіліндегі ғылыми-техникалық стиль және оған тән терминдік жүйе жаңадан ғана қалыптасып келеді деуге болады. Себебі техникалық ғылымдардың қазақ тіліндегі көрінісі немесе дерек көздері түрлі монографиялық зерттеулер, оқулықтар, мақалалар, баяндамалар, тезистер, патенттер т.б. сан-салалы еңбектер арқылы танылады. Ал бұл аталғандардың саны мен сапасы қазақ ғалымдарының әлі де болса арнайы көңіл бөлуін қажет етеді. Ана тілімізде жарық көре бастаған жаңа бастаманың нәтижесінде тілімізге жаңа ұғымдар мен техникалық жаңа атаулар келе бастағаны белгілі. Сондықтан қазақ тілінің жалпы лексикография саласының жаңа бағыты ретінде ғылыми-техникалық лексикография саласы да өз отауын құра бастады деуге болады.

Үнді-еуропа тілдерінде, соның ішінде орыс тілі бойынша аталған бағыттың өзіндік даму тарихы мен теориялық және практикалық тәжірибесі баршылық. Міне, осындай жайттарды ескере отырып, осы еңбекте ғылыми-техникалық терминологиялық сөздіктерге немесе арнаулы сөздер сөздігін құрастыруға қойылатын кейбір талаптарға қысқаша тоқталмақпыз.

Арнаулы сөздіктерді құрастырудың теориясы мен практикасы ғылыми-техникалық лексикографияның зерттеу нысаны болып саналып жүр. Мұндай арнаулы сөздіктер арқылы баспа материалдары мен әртүрлі мекеме орындарының ісқағаздарында көрініс тапқан терминдер мен әр ғылымдар салаларында

орын алатын номенклатуралық деректер жайлы түсініктеме мен анықтамалар алуға болады.

Арнаулы сөздіктер белгілі бір сала маманы мен электронды машиналар қажеттіктері үшін құрастырылады. Ғылыми сала маманына қатыстылар: біртүлді және аударма терминологиялық сөздіктер, идентификаторлар сөздігі, терминдердің жиілік сөздіктері, рубрикаторлар мен классификаторлар сөздіктері. Арнаулы сөздіктердің екінші тобына, яғни электронды машиналардың қажеттіктеріне – түрлі типті ақпаратты жүйелерге арналған ақпараттық-ізвестіру тезаурустары (*Тезаурус...*) және мәтінді автоматты өңдеу жүйелерінде қолданыс табатын машиналық сөздіктердің (*Убин*) әртүрлі типтері (негіз сөздердің, тіркестердің, қосымшалардың) жатады.

Кез келген тілдің дәстүрлі түсіндірме сөздіктеріне қарағанда ғылыми-техникалық терминологиялық сөздіктер, ең алдымен, белгілі білім аясы мен техникалық саланың мамандарына арналатыны белгілі. Терминологиялық сөздіктер, өздерінің бағышталуына байланысты – жоғары білімді мамандар мен ғылыми әдебиеттерді аударумен айналысатын кәсіптік аудармашыларға арналып құрастырылады. Жоғары дәрежелі мамандарға мұндай сөздіктер тек қажетті деген сөзді тауып алуға ғана мүмкіндік туғызып қоймай, сонымен бірге оның ұғымын анықтауға және әртүрлі басылымдарда кездесетін кейбір синонимдік сөздер мағыналарын ажыратуға да мүмкіндік туғызады.

Ғылыми мақалаларды бір тілден екінші тілге аудару кезінде идеографиялық принципте құрылған біртүлді және көптілді түсіндірмелі-аударма терминологиялық сөздіктер де аса қажет (*Берков; Убин*).

Ғылыми-техникалық терминологиялық сөздіктерді орта және жоғары техникалық оқу орындарының студенттеріне, әр салалық өндіріс орындары мен ауылшаруашылық аясындағы сан-салалы мамандарына және радио мен баспа қызметкерлеріне және т.б. арналған сөздіктер деп те ажыратуға болады. Сол сияқты, мектеп оқушыларына арналған терминологиялық сөздіктер де жеке пішінде шығарылып жүр. Бұл сөздіктердегі деректерді оқырманға ұсыну тәсілінде оқушының жас ерекшелігіне айрықша көңіл бөлу негізгі ұстаным болып табылады. Мысалы, терминологиялық сөздікте қызықтыру

элементтерімен бірге суреттердің, сызбалардың орын алуы. Соңғы кездерде ғылыми-көпшіліктік терминологиялық сөздіктермен қатар бөлек ғылыми аяның терминдік жүйесін анықтайтын арнаулы ғылыми сөздіктер де жарық көре бастады. Әдетте, мұндай сөздіктердің құрылымы әліпби тәртібінен гөрі, жүйелілік принципке негізделіп жасалады.

Қолданылу аясына қарай, тіркейтін, түгелдейтін – «узус» сөздіктерін, белгілі ережелерге бағындыруды көздейтін – регламенттеуші сөздіктерін, оқу жұмысын көздейтін – мәлімет беру, жүйелендіруші терминологиялық сөздіктер де жарық көруде (Герд, 1973; 1996).

Тіркеу міндетін атқаратын немесе «узус» атты сөздіктер қатарына терминологиялық сөздіктердің басым көпшілігінің қатысы бар. Ол сөздіктер сол кезеңге және қарастырылып отырған білім саласына тән терминологиялық стандартты анықтайды.

Регламенттеуші сөздіктер нормативті және нормативті емес деп екіге ажыратылады. Нормативті терминологиялық сөздіктерге айқын түрде мысал бола алатын сөздіктер – стандарт терминдер сөздігі мен ұсынылған (бекітілген) терминдердің сөздігі.

Стандарт терминдердің анықтамалары бойынша олар – ресми, мемлекеттік-құқықты статусқа ие болатын жүйелі типтегі терминологиялық сөздіктер. Терминологиялық стандарттардың құрамы – кіріспе, негізгі және әліпбилік көрсеткіштерінің бөліктерінен тұрады.

Кіріспе бөлікте стандарттың таратылу аясы, стандарттау нысанын анықтау және қолданылуға болмайтын терминдер, синонимдер мен аударма баламалары жайлы мәліметтер беріледі.

Стандарттың негізгі бөлімінде осы стандартқа қатысты қабылданған ұғымдар жүйесіне сәйкес орналасқан терминдер мен олардың анықтамалары қамтылады.

Терминологиялық сөздіктерге әр білім саласы бойынша Ғылыми-техникалық терминологиялық комитетінің жүйелі түрде шығарып тұратын ұсынылған (бекітілген) терминдердің жинағын да жатқызуға болады. Мұндай жинақтарда негізгі

терминдермен бірге олардың анықтамалары, синонимдері және аударма баламалары да беріледі.

Нормативті терминологиялық сөздіктерді құрастыру аса маңызды іс. Сондықтан ондай сөздік бір ғана терминдік жүйеге қатысты болуымен бірге олардағы терминдердің қолдану шарттылығы арнайы құжат арқылы бекітілуі қажет.

Арнаулы ғылым-техникалық терминологиялық сөздіктерді құрастырудың алға қойған мақсатына қарай, олар – **ақпараттаушы және жүйелендіруші** деп екіге ажыратылады.

Арнаулы ақпараттаушы сөздікке мысал ретінде ақпараттық-ізвестірудің тезаурустық түрін атауға болады. Ақпараттық-ізвестіру тезаурусы деп біртүлді қадағаланып отыратын және лексикалық бірліктерге өзгеріс енгізіп тұру сипатындағы сөздікті атайды. Бұл сөздік бір ғана тілдің лексикасына негізделіп, ақпаратты өңдеу мен ізвестіру қажеттігі үшін құрастырылады. Ақпараттық-ізвестіру тезаурусы, ол – ақпаратты автоматты өңдеу мен ізвестіруді жеңілдететін сөздік түрі. Ондай тезаурус терминдер арасындағы мағыналық, семантикалық қатынасты да толық ашуға жәрдем етеді. Ақпараттық-ізвестіру тезаурусының негізгі бірлігі – терминдер мен дескрипторлар. Оның әліпбиленген лексика-семантикалық бөлігі – бір жерге жинақталған дескрипторлық сөздік мақалалары. Әрбір дескрипторлық сөздік мақала термин-дескриптордан бастама алады да, ал одан кейін дескриптордың мақала шеңберіндегі синонимдері беріледі және негізгі дескриптормен грамматикалық немесе ассоциативті қатынаста болатын басқа да терминдер орналасады (*Тезаурус ...*).

Жүйелендіруші арнаулы сөздіктерге жататындар: *классификаторлар, рубрикаторлар, жиілікті сөздіктер және терминдік жүйе* сөздіктері. Классификаторларда реттелген түрде ғылым мен техниканың жүздеген нысан атаулары және олардың сипаттамалары беріледі. Мұнда ең алдымен техника-экономикалық көрсеткіштердің классификаторларын, өндіріспен ауылшаруашылық өнімдер атауларын және т.б. атауға болады. Осындай терминологиялық сөздіктер ішінен әсіресе рубрикатор-сөздіктер маңызды рөл атқарады. Рубрикатор-сөздіктерді сала мамандары мен жеке мекемелер үшін

ақпараттық қызмет ету жүйесіндегі әдістемелік құжат деп те атауға болады.

Рубрикаторлар, көбінде, өндірістік саланың негізгі бағыттарының жүйеленген ретін және маңызды нысандар мен процестерді қамтиды.

Салалық жиілік сөздіктердің өзіне тән маңыздылығы — кейбір бөлектелген терминдердің жиіліктерінің қамтылуы. Сөздіктің ең бір аса маңызды сипаты — терминдердің тақырыптық сипатта қамтылуы. Егер жиілік сөздік бірнеше пәндік аяны қамтитын болса, онда ол көпсалалық сөздік деп аталады. Көбіне, ғылымның дербес бөлініп шығуынан тарсалалық сөздіктер жасау қажеттігі туындайды. Мұндай сөздіктерді жасау тәжірибесінде түсіндірме терминологиялық сөздік типтерін құрастыру да аса маңызды. Түсіндірме терминологиялық сөздіктердің міндеті — нақты бір сала бойынша анықтама құралы болу және барлық терминдерді қамтумен бірге ұғымдар жүйесінің көрінісін беру. Мұндай мәліметтер мамандардың ғылыми және тәжірибелік жұмыстары үшін аса қажет деп саналады.

Терминологиялық түсіндірме сөздікте ғылым мен техниканың дамуынан пайда болатын барлық жаңа ұғымдардың терминдік көрінісін берумен қатар, ол ұғымдардың анықтамалары да қамтылуы қажет.

Терминологиялық сөздіктердегі мәліметтерді реттеуде дәстүрлі әліпбилік және ұялы әдістермен бірге статистикалық әдіс те қолданыс табады. Яғни мәліметтерді реттеуде терминнің мәтіндегі жиі немесе сирек кездесу сипаты да ескеріледі (*Алексеев*).

Өртүрлі тілдердің деректерін қамту жағынан терминологиялық сөздіктер біртілді және екітілді, үштілді және көптілді аударма сөздіктер деп бөлінеді (*Берков; Убин*). Түсіндірме сөздіктерді жасаудағы негізгі ұстаным — терминнің жаңа мағынаға ие болуы.

Терминологиялық сөздіктердің әртүрлі типтерінің негізгі дереккөзі ретінде жетекші ғалымдар мен маман-тәжірибешілердің қаламынан жарық көрген белгілі бір білім саласына қатысты жинақтап-қорытындылаушы монографиялар мен мақалалар материалдары алынады. Сонымен бірге қажетті материалдар-

дың қатарына жоғары оқу орындарының ірі оқулықтарын да жатқызуға болады. Әдетте, тек осындай оқулықтарда ғана ғылыми ұғымдар мен терминдердің аса айқын және нақты анықтамалары беріледі. Энциклопедиялар мен басқа да сөздіктерде берілген салалық терминдердің барлығы бірдей ғылыми-техникалық терминологиялық сөздіктерде орын алуы қажет.

Ғылыми-техникалық терминологиялық сөздіктердің лексикалық қорына қосымша ретінде реферативті журналдардағы кітап мазмұнының рефераттары, жаңалықтары, патенттер мен ісқағаздарының деректері де қамтылып алынады. Аса «қысылған» және шоғырланған түрде берілген мұндай мәліметтер негізгі терминдер жайлы жеткілікті түрде түсініктеме бере алады. Сол сияқты, Ресейдегі ғылыми-техникалық әдебиеттер мен құжаттарды аудару орталығының жүйелі түрде шығарып тұратын «Тетради новых терминов» атты еңбектің материалдары да терминологиялық сөздіктер үшін құнды дереккөз болып табылады.

Неше түрлі құрал-саймандарды жөндеуге қатысты және т.б. салалық, ведомстволық нұсқаулар мен нұсқаулықтарда кездесетін терминдер мен номенклатуралық белгілеулердің терминологиялық сөздік жасаудағы атқаратын рөлі де айтарлықтай мәнді.

Сонымен, сөздік түзу кезіндегі әртүрлі дереккөздерден таңдалып алынған терминдер негізінде терминологиялық сөзтізбе құрастырылады. Бұндай сөзтізбелердің реестрлік бірлігіне жататындар: сөз, сөз тіркес, қысқартулар, символдар және олардың тіркестері. Әдетте терминологиялық сөздіктің сөзтізбесі таза интуитивті тұрғыда түзіледі, яғни әрбір сала мамандарының жеке көзқарасындағы негізгі және маңызды деп аталатын терминдері таңдалып алынады. Әрине мұндай әдіс ғылыми әдебиеттерді талдауға негізделмеген, тек қана субъективті пікірлік ұстаным екені белгілі.

Ғылыми-техникалық терминологиялық сөздікті құрастырудан бұрын, маман-лексикограф ең алдымен баспаға дайындалып жатқан еңбектің хронологиялық шеңберін анықтап алуы қажет. Ғылымның жаңа салалары, мысалы, есептеу техникасы, радиофизика, биофизика үшін мұндай мәселе онша өзекті болмауы да мүмкін, себебі, ондай сөздікке енетін кез келген лексиканың уақыт мерзімі беріден басталады. Ал жаратылыс-

тану ғылымдарының — физика, механика, математика, химия, биология, география, геология және көптеген қоғамдық ғылымдар салаларында мәселе басқаша. Бұл ғылымдардың орысша мәтіндері өзінің тарихи дамуында 150–200 жылдай мезгілді қамтиды. Осы себептерден терминологиялық сөздікте көрініс табатын терминдердің хронологиялық тәртібі маңызды мәселелер қатарына жатады.

Аталған мәселені тарихи-филологиялық және тарихи-ғылыми тұрғыда кең түрде қарастыратын болсақ, ол ертеректегі мәтіндерді дұрыстап оқи білу, оның мағынасын дұрыс түсіну және бір тілден екінші тілге дұрыс аудару мәселелеріне қатысты деуге болады. Ежелгі терминдердің де қазіргі әдебиеттерде қолданыс табатынын ескерсек, аталған мәселе әр уақытта да өз құндылығын жоймайтыны белгілі.

Кейбір ғылыми кітаптар мен мақалаларда өткен кезеңдердегі ғалымдардың зерттеу нәтижелерімен пікірталас болып жатады. Мысалы, XIX–XX ғасырлардағы ғалымдардың зерттеулеріндегі ұғымдар мен концепцияларына көңіл бөліну жағы, көбінде, филология, математика, биология, химия салаларына қатысты болады. Егер мұның бәрі тек қана табиғи тіл арқылы ғана іске асатынын ескерсек, терминологиялық сөздіктерге лексикалық қорды тандап алу кезеңдік сипатта болатынына және тіл туралы ғылым тарихының өзі де ғылым тарихымен ұштасып жататынына көз жеткізуге болады.

Мәселен, орыс тіліндегі ғылыми-биологиялық мәтіндердің тарихы мынадай кезеңдерге бөлінеді:

- 1) XVIII ғ. бастап, XIX ғ. 50-жылдар аралығы — орыс тіліндегі биологиялық мәтіндердің бастапқы қалыптасу кезеңі;
- 2) XIX ғ. 50–60-шы жылдары мен XX ғ. бастапқы кезеңі (10-20-шы жылдар);
- 3) XX ғ. 10–20-шы жылдары мен бүгінгі күнге дейінгі кезең.

Биология саласы бойынша терминологиялық сөздік құру барысында XX ғ. алғашқы жылдарына дейінгі биологияға қатысты терминдерді қазіргі биология тіліне жатқызуға болмайды (Герд, 1986; 1996).

Әрине, терминологиялық сөздікті пайдалану кезінде сол сөздік ішінен термин жайлы, тіпті оның тарихына дейінгі

мәліметтер табылып жатса, әртүрлі сөздіктермен әрекет етудің қажеттігі болмас еді. Әрі тарихи тұрғыда терең хронологиялық қабатты қамтып, әрі қазіргі кездің терминдері бойынша анықтағыш рөлін атқаруға қатысты төмендегідей екі түрлі маңызды ой-пікір бар.

Біріншіден, ол – ғылыми білім саласының айтарлықтай күрделі және әрдайым құбылып тұру сипаты, ұғымдар иерархиясының күрделі эволюциясы, ұғымдардың мазмұны мен көлемінің үнемі өзгеру сипаты, ұғымдар арасындағы терминдердің тұрақсыз үлестірілуі, терминологиядағы синонимдік қатарлардың тұрақсыздығы.

Екіншіден, басқа тілдердің лексикалық параллельдерін ескере отырып, оқырманға қазіргі терминология жайлы молынан мәлімет беру жағдаяты. Сондықтан терминологияның даму тарихын қамтитын өзінше бір бөлек арнаулы сөздік болғаны жөн.

Қалай болғанда да терминологиялық сөздіктің хронологиялық төменгі шегі жайлы мәселе әлі де болса өз шешімін таппай тұр. Төменгі хронологиялық шек ХХ ғ. 20–30-шы жылдары ма немесе ХІХ ғ. мен ХХ ғ. аралығы ма? Оны тек нақты білім және ғылым салаларының негізгі кезеңдеріндегі тілдің даму сипатына қарай ғана шешуге болады (*Герд, 1973; 1996*).

Терминологиялық сөздіктердің сөзтізбесін қалыптастырғанда субъективті пікірді азайтудың бірден-бір жолы – нақты ғылым мен білім саласына тән логика-ұғымдық жүйені алдын ала модельдеу. Арнайы мәтіндерді логика-ұғымдық тұрғыда талдау тәсілі терминологиялық сөздіктерді жасаудағы алдын ала қойылатын қажетті шарт деуге болады.

Терминологиялық сөздіктердің құрылымы жайлы мынадай екі негізгі мәселені атауға болады:

1) терминологиялық сөздіктегі мәліметтерді қалыптастыру принциптері жайлы мәселе;

2) терминологиялық сөздіктегі жеке фрагменттердің құрылымы жайлы мәселе.

Бірінші мәселеге байланысты мынаны айтуға болады. Сөздік материалын ұйымдастыруға қатысты ғылыми-техникалық лексикографияда классикалық лексикографиядағы сияқты

әліпбилік, әліпбилік-ұялы, идеографиялық жалпы принциптер қарастырылады.

Терминологиялық сөздіктердің басым көпшілігі әліпби принципімен құрастырылатыны белгілі. Бұл ұстаным кез келген сөздікті пайдалану мен ондағы бірліктердің орналасу тәртібі жағынан көптен бері өзін-өзі ақтап келеді.

Осы айтылғандарға қарамай, мамандардың көптеген практикалық мұқтаждығынан аңғаратынымыз, ол терминологиялық сөздіктердегі терминдердің орналасуы тек әліпбилік немесе әліпбилік-ұялы принциптерге ғана сүйенбей, сонымен бірге логика-ұғымдық негізде, яғни ұғымдардың қатал ғылыми классификациясы бойынша идеографиялық принципте жасалуының қажеттігі. Бұл жердегі идеографикалық принципті қайсыбір ұғымды таңба арқылы белгілеп жазу деп түсіну керек. Мұндай принцип нақты ғылымның, білім саласының даму қисынына сәйкес келуімен бірге, терминді мағынасы бойынша сөздіктен тез тауып алуға да мүмкіндік жасайды. Мәселен, нақты ғылымдар мамандарына, көбінде, қайсыбір ұғымның қалай және қандай сөзбен аталатыны немесе ол сөздің синонимдерінің қандай ұғымдармен байланысы бар екендігін білу қажет болады. Маманның зерттеу тәжірибесінде және аударма практикасында белгілі бір терминді тұрпаты жағынан ұқсас болып келетін ертеректердегі зерттеулердің терминдерімен сәйкестендіру қажеттігі де туындауы мүмкін (Герд, 1996).

Идеографикалық принциппен құрастырылған терминологиялық сөздіктің негізгі құрамы – сол жағындағы терминдердің әліпби тәртібімен берілуі ғана емес, ол вербальды түрде берілген және шартты немесе формалданған тұрпаттағы ұғым болуы қажет. Аталған ұғым сөздіктің жалпы логика-ұғымдық құрылымында қатаң түрде өз орнына ие болады. Мысалы, Роженің тезаурусында (*Roget's thesaurus...*, 1978) сөздік мақаланың басы ұғымның коды арқылы (үш орынды сан) және осы ұғымды білдіретін тірек сөзбен басталады. Тірек сөзден кейін сызықша және сөз табының белгісі қойылады, сосын сол сөздің синонимдері беріледі.

Доризайфтың неміс тілінің идеографиялық сөздігінде (*Dorn Seiff, 1958, S. 211*) әрбір сөздік мақала белгілі бір тақырыптық топтағы ұғымның нөмірімен басталып, әрі қарай құрылымдық

тұрғыда бір-біріне жақын жатқан осы ұғымды білдіретін сөздің синонимдік баламалары берілген.

Ғылыми-техникалық терминологиялық сөздіктің жеке фрагменттерінің құрылымы. Лексикографияның жалпы теориясына сәйкес, сөздік мақала, көп жағдайда, мынадай бөліктерден тұрады (Герд, 1996, 297-280):

- 1) сөздік мақаланың аты;
- 2) терминдердің грамматикалық сипаттамасы;
- 3) терминнің қолданысына байланысты стильдік ренкі мен пайда болуына қатысты сипаттамасы;
- 4) терминнің семантикалық сипаттамасы.

Әліпбилік принципте түзілген терминологиялық сөздік мақаланың құрылымы түсіндірме және энциклопедиялық сөздіктердің құрылымына ұқсас болып келеді. Терминологиялық сөздіктерде сөздік мақаланың аты, әдетте, әліпби тәртібімен орналасады да, олар тек термин сөздер ғана емес, сөз тіркестер, сөздер мен символдардың комбинациясы, аббревиатуралық қысқартулар болуы мүмкін.

Әліпбилік принциппен құрылған ғылыми-техникалық терминологиялық сөздіктерде терминнің семантикалық сипаттамасын беру оның мағынасының логикалық анықтамасы арқылы немесе аударма терминологиялық сөздіктердегі баламасы арқылы іске асады. Терминнің әрі шағын, әрі дәл логикалық анықтамасының берілуі, оның семантикалық құрылымын сипаттаудың ең бір тиімді жолы деп саналады.

Терминологиялық сөздіктерде терминдердің мағынасын анықтаудың ең бір дұрыс жолы, ол әрбір ұғымның типіне қарай, бірізділікті сақтай отыра, алдын ала белгілі параметрлерге сүйену. Терминнің толық түрдегі лексикографиялық анықтамасын құруға қатысты мынадай талаптарды ұстаным еткен жөн: логика-ұғымдық сызбаға алдын ала талдау жасау, ұғымның шығу тегін анықтау, ұғымға тән маңызды белгілерді анықтау, осы ұғымның басқа ұғымдар аясы ішінен ерекше белгілерін анықтау.

Арнайы әдебиетті аудару ісімен айналысатын аудармашыға қажетті сөздің баламасын табумен қатар оның синонимдерін де іздеп табу қажет болады. Яғни терминді мазмұндық тұрғыда кеңірек танып, сөздікте көрсетілген сөз, шындығында да

түпнұсқадағы ұғымды білдіру-білдірмеуіне көз жеткізу керек болады (Берков; Убин).

Ғылыми-техникалық терминологиялық сөздіктерді құрастыруда терминнің анықтамасын тұжырымды етіп беру үшін мынадай ережелерді басшылыққа алған жөн саналады (Герд, 1996, 298-300):

1. Термин арқылы аталатын барлық ұғымдар мен нысандар анықтама ішінде қамтылуы қажет және, керісінше, анықтамада сөз болатын барлық ұғымдар мен нысандар терминнің атындағы ұғымдар мен нысандарға сәйкес келуі қажет.

2. Терминнің анықтамасында пайдаланылатын белгілер қарастырып отырған тақырыптық ая бойынша ең маңызды және мәнді болуы қажет.

3. Терминнің анықтамасында басқа бір терминнің анықтамасына сілтеме жасауға болмайды. Себебі, егер ол терминнің анықтамасында тағы да басқа бір терминнің анықта-масына жүгінсе, оның шегі табылмай, қайта оралып алғашқы терминге келуіміз де мүмкін.

4. Анықтамада пайдаланылған терминдер осы сөздік ішінде анықталған болуы немесе бұрыннан да белгілі және бірізді түсінігі болуы шарт. Егер анықтамада пайдаланылған терминдер іргелес жатқан басқа пәндерге қатысы болса немесе ол жалпы-пәндік термин болса, онда арнайы қосымшаларға сілтеме жасалғаны жөн. Басқаша жағдайда ол терминнің түсінігін қандай әдебиеттен оқып-білуге болатыны айтылуы керек.

5. Терминнің нақты мағынасын ашатын анықтама берілуі үшін оның ғылыми ұғымдар жүйесіндегі алатын орнын жете түсіну аса маңызды.

6. Терминдердің анықтамалары, көптеген жағдайда, сөздермен қатар, математикалық және химиялық формулалар мен неше түрлі безендірілулер арқылы көрініс табады.

7. Әр елдің тілдеріндегі бір ғана терминнің мазмұнына қатысты анықтамалар бір-біріне сәйкес болмауы мүмкін. Сондықтан терминнің анықтамасын берерде басқа бір сөздікке сүйенетін болсақ, оның қай елдің сөздігінен алынғаны көрсетілуі керек.

Біртілді немесе аударма терминологиялық сөздіктерде терминнің семантикалық сипаттамасы синонимия, антонимия,

текті-түрлік және ассоциативтік семантикалық қатынастар арқылы толықтырылуына әрі кеңейтілуіне болады.

Ғылыми-техникалық терминологиялық сөздікте нақты білім, ғылым, техника салаларының терминдік жүйесі көрініс табуы қажет. Ғылыми мәтіннің семантикалық мазмұнын дәл ашу, сол сала маманының қатысуынсыз мүмкін емес екендігі белгілі. Сондықтан ұғымның түсінігін сипаттау сол ғылыми сала маманының ісі де, ал оны барынша дәл және ықшам түрде тұжырымдап оқырманға жеткізу тілші-лексикографтың үлесі. Ал ғылыми-техникалық терминологиялық сөздіктің сапасы сала маманы мен тілші-лексикограф маманының бірлесе әрекет жасауының нәтижесі деуге болады.

Жоғарыда аталған арнаулы сөздіктерді құрастырудың теориясы мен практикасына байланысты тұжырымдар қазақ тілінің ғылыми-техникалық лексикографиясының дамуына оң әсерін тигізеді деген ойдамыз.

Әдебиет

Алексеев П.М. Статистическая лексикография. М., 1975.

Берков В.П. Вопросы двуязычной лексикографии. М., 1973.

Герд А.С. Основы научно-технической лексикографии. Л., 1986.

Герд А.С. Научно-техническая лексикография // Прикладное языкознание: Учебник. Отв. редактор А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. С. 287-307.

Тезаурус научно-технических терминов. М., 1972.

Убин И.И. Автоматический переводной словарь. М., 1989.

Лексикография саласындағы автоматтандыру тәсілі

Лексикографиялық нысанның ерекше түрі ретіндегі *автоматты сөздіктер* мен *автоматты терминологиялық банктердің* пайда болуы бірнеше себептермен түсіндіріледі. Біріншіден, ол ЭЕМ-ның мүмкіндіктерінің артуы болса, екіншіден, лексикографтар мен аудармашылардың еңбек өнімділігін арттыру жолдарын іздестіру қажеттігін ұғыну деуге болады.

Лексикография саласы адамның қызметтерінің ішіндегі ең бір көп еңбек сіңіруді керек ететін түріне жататыны мәлім. Сөздіктердің кейбір түрін түзу жұмысында оны сапалы аяқтап шығу үшін бірнеше онжылдықтардың өтуі мүмкін. Сондықтан лингвистикалық жұмыстарды автоматтандырудың алғашқы жетістіктерінің лексикография саласына қатысты болуы кездейсоқ жағдай емес. Неше түрлі сөзтізбелерді, үлкен көлемді сөзнұсқағыштарды, жиілік сөздіктер мен кері-әліпбилі сөздіктерді мақсатқа сай өмірге келтіру үшін оларды ЭЕМ көмегімен түзудің қажеттігі қазіргі кезде дау туғызбайды.

Автоматты сөздіктердің негізгі міндеті – адамның мәтінмен әрекет етуінде оның еңбек өнімділігін арттыру. Мысалы, лексикографиялық түсініктеме беру мақсатындағы әртүрлі тіл бірліктерін жинақтау, сақтау және өңдеу жұмыстарын автоматты сөздіктер арқылы іске асырған жөн. Яғни сөздіктерді олардың функционалдық критерийі негізінде қарастыратын болсақ, автоматты сөздіктерді айрықша сөздік түрі деуге

болады. Оның міндеті — мәтінді автоматты өңдеудің практикалық жүйесінде қолданыс табу.

Кез келген автоматты сөздіктердің міндетті түрдегі бөлігі ретінде сөзтізбе алынады. Мұндай сөзтізбе семантикалық баяндау жүйесіне немесе білім базасына енудің тетігі болумен бірге тілдік базаның бірден-бір құрастырушысы да бола алады. Мысалы, ақпараттық қызмет ету жүйелерінде түзілетін сөздіктер өз құрамында тиісті білім базасына (тезаурус, фрейм жүйелеріне) ену тетігі рөлін атқаратын дискрипторлық сөзтізбелерді сақтай алады. Ал машиналық аударма жүйелеріндегі сөзтізбелер мәтінге лексикалық және синтаксистік талдауды іске асыру үшін қажетті деген морфологиялық, синтаксистік және семантикалық сипаттамаларға ие лексикалық бірліктерді өз құрамында қорландырады. Осымен бірге мұндай жүйелер әртүрлі мағыналарды, синтаксистік және семантикалық ережелерді сипаттайтын автоматты сөздіктерді де өз құрамына енгізулері қажет. Мұндай автоматты сөздіктер машиналық аударма жүйесінің білім базасын құрайды.

Кез келген тәжірибелік жүйе жұмысы да өз құрамында білім базасына жол салатын сөзтізбе болуын қалайды.

Сөйлеу тіліндегі ақпаратты өз жадына ендіруге бейімделген жүйелердегі сөзтізбе құрамында сөзтұлғалардың жазылуы мен оның транскрипциясы да орын алады.

Түсіндірме типті біртүлді автоматты сөздік сөзтізбесі сөздік жобасының принциптеріне сүйене отырып анықталады. *Автоматты аударма сөздіктері* негізінен ғылыми-техникалық мәтіндермен әрекет етуге арналған болғандықтан, олардың көпшілігі терминологиялық лексиканы қамтулары тиіс. Ірі автоматты аударма сөздіктің нәтижелігін арттыру үшін оған тақырыптық сөздікке қатысты барлық терминологиялық лексика қамтылуы қажет. Яғни жақын тақырыптық аядағы аса жиі қолданыстағы жалпы лексика мен жалпығылымдық, жалпытехникалық лексиканың барлығы бірдей есепке алынады. Жалпылексикалық қор, ең алдымен, техника саласындағы мамандарға қажеттірек екені белгілі. Себебі, ондай мамандар, әдетте, өз саласына қатысты лексиканы жақсы меңгергенімен, шет тіліне қатысты лексиканы қажетті дәрежеде біле бермейді.

Автоматты сөздіктің міндеті ғылыми-техникалық мәтіндердің барлық түрлерімен әрекет жасау болғандықтан, оның құрамына барлық терминдер түрлері енулері қажет. Мысалы, оларға жататындар: стандарттанған және ұсынылған, сол сияқты ұсынылмаған, тұрақтанған немесе жаңадан қолданысқа енген терминдер.

Автоматты сөздіктің құрылымын анықтау жолында мынадай негізгі үш мәселенің шешімі табылғаны жөн:

- 1) сөздік мақаланың басқы бірлігінің тұрпатын анықтау;
- 2) тілдік бірлікті сипаттауға қатысты ақпарат жиынтығын анықтау;
- 3) автоматты сөздік базасындағы сөздік мақаланы құрастырудың ұтымды тәсілін таңдау.

Автоматты сөздік үшін оның сөзтізбесінің негізгі бірлігін таңдау маңызды мәселе болып саналады. Жалпы лексиканы қамтитын ірі сөздіктерде сөзтізбенің негізгі бірлігі – негіз сөз ретінде алынады. Ал әрбір сөзге қатысты сөз тіркестері, әдетте, мақала ішіндегі мағыналық көрініс ретінде немесе басқы тілдік бірліктердің тіркесімдік сипатына мысал ретінде беріледі.

Автоматты аударма сөздікке қатысты сөзтізбенің негізгі бірлігі дәстүрлі терминологиялық сөздіктердің осындай бірліктеріне жақын сипатта болулары қажет. Сөзтізбедегі лексикалық топтың 80 пайызы сөз тіркестерден тұратыны да ескерілуі қажет. Ең ірі автоматты аударма сөздік, көпшілік жағдайда, дәстүрлі терминологиялық сөздіктерге жүктелетін мәселелерді шешуге арналған терминологиялық банк рөлін атқарады. Мұндай автоматты аударма сөздіктерде дәстүрлі терминологиялық сөздіктерге қарағанда сөз тіркестерінің жоғарғы шегі анық болмауынан, оларда сөз тіркестер саны тұрақсыз болып келеді. Ірі автоматты аударма сөздіктерде дербес мақала ретінде жеке сөйлемдер мен мәтін үзінділері де есепке алынады.

Аударма сөздіктердегі сөзтізбені іріктеу мен жинақтаудың аталған түрлері, әдетте, дәстүрлі лексикографияда қолданыла бермейді. Ал автоматты сөздіктер үшін мұндай тәсіл белгілі жағдайда ұтымды болып келеді. Себебі стандартты мәтіндер бойынша жартылай автоматтанған аударма жасау тұстарында автоматты аударма сөздіктер өздерін тиімді құрал ретінде

көрсете білді. Мысалы, заң құжаттары мәтіндерінде белгілі бір мәтін бөліктері онша өзгеріссіз қайталанып отыратыны белгілі.

Қазіргі кездегі өндірістік автоматты аударма сөздіктер аса күшті ЭЕМ негізінде құрылады. Мұндай ЭЕМ бірнеше жүзмындаған көлемді сөз тіркестерінен тұратын тілдік ақпараттармен нәтижелі әрекет ете алады. Сонда да ЭЕМ әрекетінің жылдамдығы оның жадының көлеміне тікелей байланысты екендігін ескере отырып, оны үнемді пайдалануды әр уақытта ескеру қажет.

Екітілді және көптілді лексикография жағдайында сөздік бірлігін таңдау нақты тілдердің сөзжасам және тұлғажасам ерекшеліктерімен анықталады.

Айырғыш, аналитикалық, кейбір жағдайда флективті-аналитикалық тілдер типтерін өңдеуге қатысты тез әрекет ететін жүйелер құруда *«Мәтін – Автоматты сөздік»* өзара әрекеттесу сызбасы өзін-өзі ақтап жүргенін айта кету қажет. Мұндай жағдайда сөздіктегі белгілі бір сөзге тіркелген барлық ақпаратпен танысу үшін синтаксистік деңгейге дейінгі морфологиялық талдау алгоритмінің көмегіне сүйенбеуге де болады. Сонымен бірге автоматты сөздіктегі сөзтұлғаларды олардың мәтіндегі «егіздерімен» теңдестіру аса үлкен жылдамдықпен іске асады. Ал синтетикалық, флективтік және агглюнативтік тілдердің автоматты сөздіктерінің бас бірлігі ретінде сөзтұлға алынатын болса, сөздік көлемі кенет ұлғаяды да мәтін бойындағы сөзтұлғаны автоматты сөздіктегі сөзтұлғамен теңдестіру жылдамдығы, керісінше, азаяды. Мысалы, орыс тіліндегі зат есім сөз 12-ге дейін, ал етістік сөз, егер түр корреляциясын есепке алатын болсақ, 188 рет тұлға өзгерісіне ұшырайды екен. Түркі тілдерінде есім түріндегі тілдік бірлік саны, теориялық жағынан алғанда, тұлға өзгерісіне шексіздікке дейін ұшырауы мүмкін.

Осы айтылғандарға байланысты мұндай парадигмаларды компрессиялау (қысу) қажеттігі туындайды.

Бұл мәселенің шешімі екі түрлі тәсілмен іске асуы мүмкін:

- 1) лингвистикалық (алгоритмдік);
- 2) техникалық.

Лингвистикалық тәсілде негізгі назар сөздіктің бірлігін қалыптастыруға бөлінеді және оның эталонын, үлгісін және ең

ыңғайлы баламасын табуға әрекет етіледі. Синтетикалық, флективтік және агглюнативтік тілдер үшін неше түрлі компрессиялау тәсілдері қолданылып жүр. Олардың ішінен ең нәтижелі деген әдіс ретінде автоматты сөздікті негіз сөздердің сөздігі ретінде қалыптастыру болып табылады. Мұндай автоматты сөздіктерді, әдетте, тілдік бірліктің екі түрі жайлы сөз етуге болады: *негіз сөз және ерекше тұлға*. Басқаша айтқанда олар сөз негізінің алмасуы арқылы немесе супплетивті (әртүрлі негізден жасалған бір сөздің тұлғалары) тәсіл арқылы жасалған сөзтұлғалар. Ерекше тұлға бірліктер автоматты сөздіктің құрамына негіз сөздермен бірдей дәрежеде енгізіледі.

Сонымен бірге бірқатар тілдер үшін парадигма ретінде тек жалғыз негіз сөзді ғана енгізу жеткілікті бола бермейді. Бұл жағдай тек ерекше тұлғаларды қолдану кезінде де жеткіліксіз.

Дамыған флексиясы бар орыс тілінің лексикалық жүйесін модельдеу тұрғысынан негіз сөздер сөздігі қолданылады. Мысалы, А.И.Герцен атындағы РМПУ-дың инженерлік зертханасы зерттеп дайындаған және СИЛОД жүйесінің өзегі болып саналатын орыс тілінің «МАРС» автоматты сөздігі негіз сөздер мен типтік парадигмалар сөздігі негізінде құрылған. Ал GETA жүйесін сөздікпен жабдықтау үшін орыс тілінің мынадай жеті сөздігін іске қосу алдын ала қарастырылған: алғашқы екеуі – екі негіз сөздер сөздігі, келесі екеуі – өзгеріске ұшырамайтын сөздер мен ерекше тұлғалардың, тағы да екеуі – жұрнақ пен жалғаулардың сөздігі және соңғы бір сөздік – машиналық тіркесім сөздігі.

Сөздік мақаланың құрамы әртүрлі автоматты аударма сөздіктерде өзгеше және олардың зоналық айырмашылығы 10-нан 99-ға дейін құбылып отырады. Аталған зоналарда әр типті лингвистикалық және экстралингвистикалық ақпараттар тіркеледі.

Әр автоматты аударма сөздік өзінің сөздік мақаласының ішінде белгілі реттілікпен мынадай мәліметтерді қамтиды: басқы сөз, оның бір немесе бірнеше шеттілдік баламалары, басқы сөздің тақырыптық қатыстылығы және оның шеттілдік баламалары, грамматикалық ақпарат, формалданбаған түсініктемелер немесе стандарттанған дефинициялар, мәнмәтіндер, лексикалық тіркесім, басқы сөздің семантикалық иерархиялық

байласымы, сапалық коды (сенімдік индексі), қысқартулар, стильдік мәнері (баяуы), лингвогеографиялық шектелуі, тірек сөздер (егер басқы лексикалық бірлік сөз тіркес болса), дерек көздер, сөздік мақаланың авторы, шығу мезгілі, әртүрлі қызметтік ақпараттар.

Сөздік мақаланың құрылымындағы семантикалық, синтаксистік және морфологиялық қатынастар арқылы мақаланың өзіне тән жазылу ерекшелігі анықталады.

Деректер базасында орын алатын әрбір лексикалық бірлік морфологиялық, синтаксистік, семантикалық және функционалдық деңгейлерде сипаттамалар жиынтығы ретінде көрініс табады. Синтаксистік және морфологиялық сипаттамалар мәтін бойында кездесетін сөзтұлғаның мағынасын анықтау үшін және оның сөйлемдегі синтаксистік қызметін белгілеу мақсатында қажет. Семантикалық сипаттамалар лексикалық бірліктің мағыналық жағын баян етеді де, ал функционалдық сипаттамалар бірліктің барлық құжат құрылымындағы потенциалды рөлін анықтайды.

Қазіргі кездегі мәтінді автоматты өңдеу жүйелерінде лексикалық бірліктің жоғарыдағы сипаттамаларына тиісті орын беру мақсатында сөздік мақалалардың төмендегідей үш түрі қолданыс табады.

1. Көлемі алдын ала белгіленген, бірыңғай катал құрылымды сөздік мақала. Бұл құрылым лексикалық бірлік тобын сипаттайтын барлық өлшемдерді мақала ішінде қамту әрекетін алдын ала қарастырады. Мұндай тәсіл лексикалық бірліктерді баяндаудың барлығына бірдей қатысты. Оған лексикалық бірліктің семантикалық және функционалдық өлшемдері де қоса беріледі. Аталған тәсіл бойынша лексикалық бірлікті баяндау тек синтаксистік ақпаратпен ғана шектелуі мүмкін. Бұл жағдайда лингвистикалық ақпараттық базаға семантикалық ақпарат жайлы мәлімет жеке сөздік ретінде енгізіледі.

Осындай мақала құрылымында ақпараттың барлық типтері мен олардың байттық көлемдік шегі алдын ала анықталып, белгіленеді. Егер ақпараттың нақты типі анықталмаса, оның орны «бос» күйінде сақталуы қажет. Мұндай катал құрылымды сөздік мақалаға мысал ретінде жалпы көлемі 225 байт алатын

SYSTRAN жүйесіндегі мақаланы атауға болады. Бұл жүйенің автоматты сөздігі негіз сөздер сөздігі ретінде құрылған. Әрбір сөздік мақала өз бойында мынадай мәліметтерді сақтайды:

Ү — лингвистикалық ақпараттық базаға енгізу кезіндегі лексикалық бірлікті баяндау тәсілі (жаңадан қосу, түзету, жою);

А — базалық машиналық негізді баяндау (описание);

В — септік жалғаулар жайлы морфологиялық ақпарат, көптік жалғаулардың түр-тұрпаты және т.б.;

С — меңгеруге қатысты мәліметі бар синтаксистік кодтар;

Д — семантикалық кодтар;

Е — сөз алдында келетін көмекші сөздердің кодтары (барлық сөз таптары үшін бұл орын бос болуы керек);

Ғ — балама ерекшеліктерін баяндауды қамтитын негізгі аудармалар.

Сөздік мақаланы құрудың мұндай әдісі ондағы ақпаратты оңай жолмен толықтыру мен үнемі өзгертіп отыруға барынша мүмкіндік тудырады. Себебі автоматты сөздікте әрбір ақпараттың алдын ала белгіленген өз орны болғанымен, оларда ең басында ескерілмеген жаңаша типті ақпаратты базаға енгізу сөздік мақаланы, сөздікті және базаның өзін түбегейлі өзгертуге әкеп соғады. Сондықтан мәтінді автоматты өңдеу жүйелерінде лексикалық бірліктің барлық сипаттамалары толық қамтылғаны жөн болады. Әрине мұндай талап, әсіресе өндірістік жүйелерде ЭЕМ-ның қомақты жадын алуымен бірге автоматты сөздікті орналастырудың қиындық тұстарына әкеп соғады. Солай бола тұра бұл әдіс, көбінде, ақпаратты екінші рет қайталаудан сақтандырады.

2. Оң жақ шегі жылжымалы болып келетін сөздік мақала. Мұндай мақаланың баяндалуы — «таңба — мағына» принципі бойынша жүзеге асады. Аталған тәсіл бойынша мақаланы баяндауды қалыптастыру бірізді сызба арқылы жүргізіледі және лексикалық бірлікке қатысты барлық ақпаратты зоналарға үлестіру жолдары қарастырылады. Ондай зоналардың саны нақты мәліметтердің типіне байланысты болып келеді. Ал лингвистикалық бірлікті баяндау үшін, сөзді топтастыру тұрғысынан алғанда, лексика-грамматикалық, морфологиялық, семантика-синтаксистік, тұрпаттық, функционалдық және т.б. көзқарастағы сипаттамаларға сүйенеді. Әрбір

зонада лексикалық бірліктің қандай топқа қатыстылығына қарай, белгілер саны жекеленіп алынады. Мысалы, зат есімді актуалдаудың функциясын атқаратын детерминативтер класы осы кластың нақты бірлігін баяндайтын мынадай белгілер жиынтығымен сипатталады: сөздің ішкі топтары, өзгеріп түрленуі, орналасу қалпы және т.б.

Сөздік мақаланың «оң жақ шегінің жылжымалы» болуы мақаланы тек ЭЕМ-ға (компьютерге) дайындау кезіне ғана сәйкес келеді. Ал компьютердің жадына енгізу тұсында мақаланың баяндалуы, алғашқыдай қатаң құрылымға ауыстырылуы керек.

3. Иерархиялық құрылымдағы сөздік мақала. Бұл тәсіл ақпарат бөліктерін жеке файлдар (массивтер) ретінде бөліктеп шығаруды алдын ала көздейді. Мұндай жағдайда автоматты сөздік, ең аз дегенде, екіге ажыратылады: міндетті түрдегі тіркелген ақпараты бар сөздікке ену тізімі және мағыналар сөздігі.

Лексикалық бірліктің мәтінді автоматты өңдеу жүйесінде қолданылу мүмкіндігін ескере отырып, оның барлық сипаттамаларын иерархиялық топтастыру алдын ала жүргізіледі, яғни иерархиялық жоғарғы және төменгі деңгейлері анықталады. Мәтінді автоматты өңдеудің барлық түрлерінде тікелей қолданылатын жоғарғы деңгейдегі сипаттамалар мақала құрылымына тікелей енгізіледі, ал барлық басқа сипаттамалар (әдетте семантикалық) жеке файл түрінде көрініс табады. Мұндай жағдайда сөздік мақалада жекешеленген файлдарға арнайы сілтеме беріледі.

Базалық вариантта сөздік мақала құрылымының (базаға «енетін») сөздік құрамына жататындар: басқы лексикалық бірлік (негіз, сөзтұлға немесе сөз тіркесі); ендірілетін тілдің типологиялық ерекшелігіне, грамматикасына байланысты лексика-синтаксистік код; сөздің әртүрлі заттық аяда атқаратын қызметінің ерекшелігін анықтайтын ішкі тіл (подъязык) коды; аударма жайлы ақпарат.

Баяндаудың жоғарғы деңгейі, яғни сөзтізбе өз құрамында лексикалық бірлікті және сөйлемге алғашқы синтаксистік (синтагмалық) талдау жүргізуге қажетті морфологиялық сипаттамаларды қамтиды. Мысалы, сұрақ-жауап жүйесінде

баяндаудың жоғарғы деңгейі – әрқайсысына синтаксистік және морфологиялық ақпарат тіркелген лексикалық бірліктің өзінен ғана тұруы мүмкін. Бұл деңгей анықтағыш деп аталып, іштей екі бөліктен тұрады (Беляева, Герд, Убин, 324-с).

Енді автоматты аударма сөздік арқылы орындалатын лингвистикалық баяндаулардың кейбір ерекшеліктеріне тоқталайық.

Бұл мәліметтер Бүкілодақтық аударма орталығында (ВЦП) зерттеліп, сөздік мақаланың құрылымына енгізуді көздейді. Сөздік мақала 13 зонаға бөлініп, әрбір зона қатаң түрде белгілі ақпаратпен өзіндік ереже бойынша толтырылады.

1-зона. Сөздік мақаланың атын анықтайтын тірек зона. Бұл зона бойынша сөздік мақалаға «ену» әрекеті жүзеге асады және оны қалыптастыру мен іздестіру жұмыстары орындалады. Мақаланың аты ретінде: морфема, жеке сөз, канондық тұрпаттағы сөз тіркесі немесе мәтін үзіндісі алынуы мүмкін. Толық омонимдер үшін рет саны тіркелген әртүрлі сөздік мақалалары құрылады.

2-зона. Сөздік мақаланың атынан кейін оның лексика-грамматикалық класын анықтайтын 2-зона орналасады. Бір элементті лексикалық бірліктер сөз таптарына қатыстылығына қарай топ-топқа бөлінеді. Бұл топтардың ішінде лексикалық бірліктер бірнеше категориялық белгілері бойынша тағы да ішкі кішігірім топтарға ажыратылады. Сөз тіркестері өздерінің сөйлем ішіндегі атқаратын қызметіне қарай типтік түрлерге бөлінеді.

2а-зона. Орыс тіліндегі аударма эквиваленттің лексика-грамматикалық класын анықтайтын зона. Тікелей орыс тіліндегі аударма эквивалентінен кейін оның лексика-грамматикалық класы жазылады. Орыс тіліндегі аударма эквивалентінің сөздік мақаланың құрамына енуі сөздік мақаланы «керісінше» орын ауыстыру үшін қажет, яғни *шет тілі – орысша* сөздік мақаладан автоматты түрде, керісінше, *орысша – шет тілі* сөздігінің мақалаға ауыстыру мақсаты үшін.

3-зона. Морфо-грамматикалық ақпарат зонасы. Бұл зонада басқы бірліктің барлық сөз өзгерту парадигмаларын құру әрекеті орындалады да, соның нәтижесінде кез келген мәтіндік тұрпатты іздестіру мүмкіндігі туады.

Морфо-грамматикалық ақпарат зонасында басқы лексикалық бірліктерді қысқартып жазу варианттары ескерілетін ішкі зона қарастырылған. Мұндай мүмкіндік жеке тұрған сөздерге де, сөз тіркестеріне де қатысы бар.

3а-зона. Орыс тіліндегі аударма эквивалентіндегі қысқартулар зонасы. 2а-зонадан кейін 3а-зона енгізіледі. Яғни лексика-грамматикалық кластан соң орыс тіліндегі аударма эквивалентінің қысқартулар зонасы орналасады.

4-зона. Тақырыптық рубрика (сөзбасы) мен стиль зонасы. Бұл зонада тақырыптық және стильдік белгілер беріледі. Ондай белгілер көмегімен басқы бірліктің осы мағынадағы қолдану аясы мен шарты нақтыланады. Осы зонада шет тілдерінен алынған кірме лексиканың дереккөзіне сілтеме жасалуы мүмкін. Осындай белгілер мен таңбалардың типтері негізінде белгілі стилистикалық ерекшелігі бар монотематикалық сөздіктерді немесе сөзтізбелерді автоматты түрде қалыптастыруға болады.

5-зона. Сенімділік индексінің зонасы. Әрбір орыс тіліндегі аударма эквивалентінің алдында осы аударма эквивалентінің (баламасының) көпшілік мақұлдаған сенімділік индексі тұрады. Бұл индекс лексикалық бірлік алынған лексикографиялық дереккөздің сенімділігімен анықталады. Сенімділік дәрежесін белгілеу үшін орыс тіліндегі аударма эквиваленті жүйесінде мынадай индекстер қабылданған:

А – қайсыбір стандартта (ГОСТ, ОСТ, халықаралық стандарт) бекітілген немесе оның қолданысын беделді халықаралық мекеме ұсынған аударма эквиваленті (баламасы).

Б – аударма эквиваленті сенімді және апробациядан өткен сөздіктерде берілген.

В – аударма эквиваленті жаңа терминдер дәптерінде, терминологияның экспресс-шығарымдарында және т.б. берілген.

Г – аударма эквиваленті аударылған журналдарда, баспадан шықпаған аудармаларда кездескен немесе сөздік мақаланы құрастырушының ұйғарымымен алынған.

6-зона. Орыс тіліндегі аударма эквивалент зонасы. Бұл зона канондық тұрпатта жазылған сөз немесе сөз тіркестерінен тұрады. Барлық орыс тіліндегі аударма эквиваленттер қатар

сандарымен беріледі. Синонимдік қатарды құрайтын орыс тіліндегі аударма эквиваленттері бірдей нөмірмен жазылады, ал басқы бірліктің әртүрлі мағыналарын білдіретін лексикалық бірліктің аударма баламалары әртүрлі нөмірленеді.

Дәстүрлі лексикографияда *баламасыз лексика* деген ұғым бар. Мұндай лексикаға басқа тілдерге аударылмайтын лингвистикалық бірлікті жатқызады. Баламасыз лексика басқа елдерде жоқ ұғымды білдіретін жалпы қолданыстағы лексика немесе басқа тілдерде әлі қолданыла қоймаған жаңа терминдік лексика болуы мүмкін.

Егер жаңа термин өзінің аударма баламасының мағынасын ашып жатса, оны орыс тіліндегі аударма эквивалентіне қосып жазуға болады. Бұдан кейін сөздік мақала бір қалыпқа түседі.

7-зона. Түсіндіргіш белгілер (таңбалар) зонасы. Түсіндіргіш белгілер немесе таңбалар лексикалық бірліктердің мағынасын нақтылау үшін және олардың мәтіндегі қолдану ережелерін сипаттау үшін қажет. Бұлардың ақпараттағы аударма баламасын дұрыс таңдап алуға да жәрдемі бар. Сөздік мақалаларын «кері қарай аудару» жағдайында, яғни **орысша-шетелдік** нұсқаны автоматты түрде қалыптастыру жағдайында түсіндіргіш таңбалар шетел лексикалық бірлік рөліндегі аударма баламасында да өз орнын сақтайды. Түсіндіргіш таңбалар лексикалық бірлікке қойылатын талаптарға сай, семантикалық терминдер негізінде де көрініс табуы мүмкін. Яғни орыс лексемасына сәйкес келетін бас актанттың орнында қандай семантикалық топтың өкілі тұратынына нұсқайды.

Семантикалық таңбалар басқы лексикалық бірліктің және оның аударма баламасының семантикалық сипатының да көрінісі болуы мүмкін.

Лексикалық таңбалар орыс тіліндегі аударма эквивалентті және басқы лексикалық бірліктің лексикалық тіркесімдік жағдайын нақты лексемалар тұрғысынан анықтай алады.

Синтаксистік таңбалар орыс тіліндегі аударма эквиваленттерді басқарудың синтаксистік моделін баяндайды. Бұл жағдай синтаксистік орын мен басқы бірліктің осы мағынадағы грамматикалық әрекетінің ерекшеліктері тұрғысынан сипатталады.

8-зона. Түсінік беру зонасы. Терминологиялық бірліктердің мағынасын баяндау үшін түсініктеме беру тәсілін қолдануға болады. Автоматты аудару сөздігі жүйесінде омонимдік және көпмағыналық құбылыстардың шекарасын ажыратуда түсініктеме беру ең бір ұтымды тәсіл болып табылады.

9-зона. Басқы бірліктің қолдану жағдаятына мысал келтіру зонасы немесе тілдік бірліктің қолдану мәнмәтіндерінің зонасы. Басқы лексикалық бірліктің белгілі мағынадағы қолдану мүмкіндіктері орыс тіліндегі аударма эквивалентінде тиісті мысалдармен қамтамасыз етілуі мүмкін.

10-зона. Фразеологиялық зона. Басқы сөзді немесе сөз тіркесін, олардың лексикалық жүйеден алатын орны тұрғысынан сипаттайтын ақпаратты жазуға арналған ішкі зона (подзона) бөлініп алынады.

Мәтінде әртүрлі көрініс табатын лингвистикалық бірліктер жиі кездесетіні белгілі. Мұндай жағдайда лингвистикалық бірліктердің жазылу түрінің санына қарай, сөздік мақала саны құрастырылады.

11-зона. Сөздік мақаланы құрастырушының немесе редактордың аты-жөнін тіркейтін зона: фамилиясы, аты және әкесінің аты.

12-зона. Жазба материалдың немесе пайдаланылған деректердің аттарын тіркейтін зона. Әрбір аударма баламасына сөздік мақалада оның қандай дереккөзден алынғаны көрсетіледі.

13-зона. Сөздік мақаланы құрастыру немесе редакциялау мезгілін тіркейтін зона: жылы, айы, күні.

Жоғарыда сипатталған 13 зонаға қатысты нақты мысалдармен Л.Н.Беляева, А.С.Герд, И.И.Убиндердің «Автоматизация в лексикграфии» атты мақаласы бойынша толығырақ танысуға болады (*Прик. язык, с. 318-333*).

Сөйлем түсінігінің синтаксистік, семантикалық көпмағыналығымен “күресу” үшін және нөлдік деңгейдегі семантиканы ұсыну қажеттігінен көбінде сөз орамдарының автоматты сөздігі қолданылады. Мұндай сөздік бойынша мәтіндегі сөздер тізбектерін бөліп алып, оларды екінші тілге аударуға мүмкіндік туындайды. Сөздер тізбегінің мағынасы әрбір элементтің мағыналарының жиынтығы еместігі ескеріледі.

Лингвистикалық ақпараттар базасы құрамына аталған сөз орамдарының автоматты сөздігі дараланып немесе автоматты сөздікпен бірге енуі мүмкін. Бұл сөздіктердің (сөз орамы сөздігінің) қажеттігі қазіргі кездегі техникалық құралдардағы тілдік деректердің жеке сөздермен емес, сөз тіркестерімен берілетіндігімен түсіндіріледі.

Тілаларлық идиомалық сөздердің машиналық аударма кезінде мәтін мағынасын тануға қиындық туғызатыны белгілі. Бұл қиындықты семантикалық талдау алгоритмін қолданып шешу де оңай емес. Сондықтан идиомаларды дұрыс аудару үшін сөз орамдары сөздігін машиналық аударма жүйесіне енгізу қажеттігі туындайды.

Сөз орамдарының сөздігін құрастыруда сегменттер шекарасын бөліп алу үшін статистикалық критерийлер ескеріледі. Мысалы, сөз орамдарының автоматты сөздікке ғылыми-техникалық мәтіндер бойынша жасалған әртүрлі жиілік сөздіктердегі аса жиі қолданылатын сөз тіркестері енгізіледі. Сонымен бірге ол сөздікке сөз тіркестерінің аударылу критерийлері де қоса беріледі. Басқаша айтқанда, сөз орамдарының автоматты сөздігінде бастапқы мәтін үзіндісінің тікелей аудармасымен сәйкес келе бермейтін мәтін үзінділері де қамтылады.

Тілдің семантикалық жағын баянды ететін автоматты сөздік дербес модуль ретінде бірнеше тәсілдермен құрылады:

а) белгілі ережелер негізінде біріктірілген семантикалық белгілердің жиынтығы түрінде;

ә) кеңейтілген немесе ассоцияланған (байласымды) семантикалық желілер ретінде.

Әдетте, күрделі автоматты аударма жүйелерінде бірнеше тақырыптық аяның лексикалық топтары қамтылуы керек. Себебі әр тақырыпқа қатысты лексиканың бір-бірімен қиылысу мәселесінің шешімін табу автоматты аударма жүйесінде аса маңызды іс болып саналады. Өндірістік автоматты аударма жүйесін құрастырушылар аталған мәселеге үлкен мән беріп, осы мақсаттың шешімін табуға арнаулы лексиканы тақырып бойынша топтастыру әрекетін жүзеге асыратын жүйелерді іске қосуда. Автоматты аударма жүйесіндегі лексиканы тақырыптық топтауға негіз ретінде ғылыми-техникалық ақпараттың

Мемлекеттік автоматтандырылған жүйесі орталығының басқы сөздері (рубрикаторлары) алынған.

Машиналық аударманың автоматты сөздігінің қатан түрде нақты тақырыптық аяға бағышталуы лексикалық талдау деңгейінде тілдік бірліктердің көпмағыналығын ажыратуға және терминдерді аудару мәселесін стандарттауға мүмкіндік тудырады. Нақты тақырыптық аяға бағышталу автоматты сөздіктің кез келген жүйесі үшін бірден-бір қажет. Ал егер автоматты сөздіктің көпаспектілі жүйесін құру керек болса, онда тілдік бірліктің нақты тақырыптық аяға қатыстылығын ажырату қажеттігінен сөздік мақаланың құрылымына арнайы шартты белгі-код енгізіледі. Сол код арқылы лексикалық талдау кезінде тілдік бірліктің аударма баламасы таңдалады.

Көптілдік автоматты аударма жүйесін құрастырудың негізгі қағидасы ретінде автоматты сөздіктегі әр тілге қатысты тілдік бірліктердің өзара байласымдық сипатын анықтау тәсілі алынады. Өндірістік автоматты аударма жүйесінде тілдердің санына қатыссыз лексикааралық байланысты анықтаудың негізгі үш түрлі тәсілі қолданылады:

1. Әрбір лексикалық бірлік үшін басқа тілдерге қатыссыз, тек нақты осы тілдің өз мүмкіндігі (құралдары) негізінде ғана баяндалады (басқы сөз жайлы мәліметтер, грамматикалық ақпарат, лингвистикалық және т.б. да ақпараттар);

2. Аударма баламасы ретінде жүретін екі түрлі тілдердің тілдік бірліктеріне ортақ болып келетін сөздік мақала құрастырылады.

3. Автоматты сөздікке енгізілген барлық тілдердің аударма баламаларына ортақ болып келетін сөздік мақала құрастырылады.

Сөздіктің «керіқайтқыштық» қасиеті деген ұғымды тұтынушының қалауына қарай, екі түрлі тілдің қайсысын болса да автоматты жүйеге енетін немесе одан шығатын тіл деп санап, орындарын еркін ауыстыруға болатын мүмкіндігі деп түсіну қажет.

Көптілдік автоматты аударма жүйесінде керіқайтқыштық мәселесі екі түрлі деңгейде қарастырылады – тілдер деңгейінде және жеке лингвистикалық бірлік деңгейінде. Ондай жүйедегі керіқайтқыштықтың тілдік деңгейдегі түрінде сөздікке енген

тілдердің қайсысын болса да *сұраныс тілі* ретінде және *жауап тілі* ретінде қарастыру мүмкіндігі бар. Ал лингвистикалық бірлік деңгейіндегі керіқайтқыштықты (лексикалық керіқайтқыштықты) бір ғана лексикалық бірлікті сұраныс және жауап ретінде қолдану мүмкіндігі деп ұғыну қажет.

Барлық ірі автоматты аударма жүйелері жарым-жартылай немесе толық түрдегі керіқайтқыштық қасиетке ие болып келеді. Ал автоматты аударма жүйесінің керіқайтқыштық қасиетке ие болмауы, оның аса үлкен кемшілігі деп саналады. Сонымен бірге олардың аударма терминологиялық сөздіктермен де көптеген ортақ қасиеттерге ие екендігі белгілі.

Автоматты аударма жүйесінің керіқайтқыштық қасиеті оны дәстүрлі түрдегі аударма лексикографиясындағы кейбір автоматтандыру жұмыстарында қолдануға мүмкіндік тудырады. Аударма сөздігін түзудегі уақытты үнемдеу сипаты ондағы тілдерді, сөздік мақалалардың жүйеге енетін және одан шығатын тілдері ретінде, автоматты түрде орындарын ауыстырып тұруына қатысты болады.

Барлық ірі автоматты аударма жүйелерінің негізгі ерекшелігі олардың көптілділігі. Мысалы, 20 жылдай бойы тәжірибеге енген LEXIS атты автоматты сөздік қазіргі кезде мынадай лексикалық топтарды өз құрамында сақтайды: неміс, ағылшын, француз, орыс, итальян, дат және португал тілдері. Сол сияқты, EURO DICAUTOM автоматты аударма жүйесі алты түрлі тілді қамтиды: ағылшын, дат, итальян, неміс, нидерланд және француз тілдері. Ал қазіргі кезде TFAM автоматты аударма жүйесі тоғыз түрлі тілді қамтып, олардың 2 млн. көлемді тілдік бірліктерімен аударма жұмыстарын атқара алады. Кезіндегі Бүкілодақтық аударма орталығының (ВЦП) автоматты сөздігі және Дрездендегі техникалық университеттің терминологиялық банкі өз құрамында төрт түрлі тілді қамтиды: ағылшын, неміс, француз және орыс тілдері.

Дәстүрлі аударма сөздіктеріне қарағанда автоматты сөздік динамикалық қасиетке ие, яғни олар сөздіктің сөзтізбесі мен сөздік мақаланың көлемділігіне қарамай, оларға қажетті түзетулерді тез және еш қиындықсыз енгізе алады.

Автоматты сөздікті құрастыру – ғалымдардың лексикалық топты таңдау жұмысымен тұрақты айналысуларының арқасын-

да, сөздіктің халық алдында қызмет ету кезеңдерінің бәрінде де өз жалғасын тауып отырады. Қажетсіз деген сөздерді жүйеден шығарып тастап және оны жаңа лексикамен толықтырып, сөздік мақаланың кез келген тұстарын әрі түзетіп, әрі толықтырып отыру мүмкіндігі тұрақты түрде орын алады. Құрамының айнымалылығы мен сөздіктің қызметі кезіндегі түрлі түзетулерді қабылдау мүмкіндігі – машиналық сөздіктердің тұрақты сипаттағы және дәстүрлі сөздіктерден айырым табатын бірден-бір сипаттамасы.

Соңғы жылдары аса кең орын ала бастаған сөздік – ЭЕМ арқылы құрастырылатын конкорданс-сөздіктер. Қазіргі кезде Шекспир, Ронсар, Пушкин және т.б. классиктердің шығармалары бойынша іргелі конкорданс-сөздіктермен компьютер арқылы танысуға мүмкіндік туындап отыр.

Сөзнұсқағыш сөздік дегеніміз әліпби ретімен реттеліп жазылған және сөздердің қандай мәтіндерде қолданылғанына сілтеме жасалатын сөз немесе сөзтұлғалар сөздігі. Конкорданс-сөздік аталған мәліметтерді өз бойында ұстайды және сонымен бірге ол сөздікте әрбір сөзге қатысты мәнмәтіндер де бірге қамтылып отырады. Түсіндірме сөздіктерге қарағанда конкорданс-сөздіктерде сөз мағынасын анықтауға қатысты түсініктеме беру және қажетті деген мысалдармен жабдықтау қарастырылмайды.

ЭЕМ көмегімен түзілген конкорданс лексикографты көптеген әрекеттерден босатады. Мысалы, оған жататындар: сөзтізбені тандау мен құрастыру, мәнмәтін кәртiшкілер қорын жинақтау және т.б. Егер кез келген конкорданс-сөздігіне лексикограф күшімен қажетті деген толықтырулар жасалынатын болса, ондай сөздік түсіндірме сөздіктердің түр-түрін құрастыруға негіз бола алады.

Конкорданстарды құрастыру жұмыстарын талдай келе, мынадай тұжырым жасауға болады. Қазіргі кезде ең тиімдісі – арнайы және толық түрде автоматтанған сөздік-анықтағыш қызметін атқаратын «сұрақ-жауап» ұстанымында құрылған конкорданс типтес сөздікті жобалау және оны іске қосу. Осындай жұмыстың негізінде тұтынушының өзіне қажетті деген шығармалар мәтінін компьютер жадына бір-ақ рет енгізіп, одан кейін өз қалауынша автоматтанған жүйе ретіндегі конкордансқа

өз қалауынша “енуіне” мүмкіндігі бар. «Деректер банкісі» деп аталатын мұндай жүйелер мәтін ішіндегі ақпараттың тек қана дәлдігін, сенімділігін және дұрыстығын тексеріп қоймайды және сонымен бірге тиісті сөздің барлық қолданысы жайлы, әрбір сөзге (сөзтұлғаға) қатысты барлық мәнмәтін түрлері жайлы және шығармалар жиынтығы жайлы толық мәліметтер бере алады.

Мұндай автоматтанған сөздік-анықтағыш қызметі ежелгі жазба ескерткіштерге, жеке авторлардың классикалық әдеби шығармаларына, арнайы терминологиялық жүйе көрінісіне, әр тілде жазылған мәтіндерге барынша бейімделе алады.

Сөздік-анықтағыш қызметін атқаратын автоматтанған лингвистикалық деректер қорын құрудың өзектілігі еш күмән туғызбайды және мұндай мүмкіндіктің “ауылы” да аса алыс емес.

Сонымен, сөздік-анықтағыш жүйесін құру болсын немесе конкорданстар жүйесін құру болсын бұлардың бәрі де өзекті жалпы-филологиялық мәселе екені айқын. Мұндай жүйелерді ойдағыдай жүзеге асыру үшін тілшілердің, әдебиетшілердің, математика және қолданбалы лингвистика мамандарының тығыз түрде одақтасып, бірлесе жұмыс атқарулары қажет екені анық.

Терминологиялық жұмыстарды механикаландыру мен автоматтандыру қызметін ғылыми-техникалық әдебиеттер мен құжаттар мәтіндерін дәстүрлі тәсілмен өңдеу әрекеттерімен бірге, оларды жүзеге асыру бірнеше функционалды элементтерден тұратын терминологиялық банк құруды, яғни автоматты жүйе құруды қамтамасыз еткен болар еді. Бұл жүйелердің негізінде барынша аз мезгілде, оңай жолмен және ең төмен қаражатпен терминологиялық деректерді жинақтау, сақтау және тарату жұмыстарын іске асыруға болады.

Соңғы кезде Ресей мен бірқатар басқа да шетелдерде (Франция, Швеция, Бельгия, Канада) әртүрлі білім салалары бойынша деректердің құрама терминологиялық деректер банкін құруға байланысты жұмыстар кең түрде өрістей бастады. Келесі кезекте тұрған жұмыс түрі — банктар арасындағы телебайланыс қызметін іске қосу әрекеті.

Мәселен, Ресейдегі автоматтанған ақпараттық-терминологиялық жүйенің қызмет атқаруының негізінде — ғылым мен техниканың әртүрлі аясындағы терминдерді қамтитын «терминдердің анықтағыш банкі» құрастыру жүргізілуде. Бұл банкке қажетті ақпараттық дереккөзінің рөлін ғылыми-техникалық терминдер мен анықтамаларға арналған мемлекеттік және салалық стандарттар, терминологиялық халықаралық стандарттар, ұсынылған терминдердің жинақтары атқарады.

Негізінде, терминологиялық деректер банкі, бір жағынан, ұғымдар мен терминдер туралы және олардың қалай таңбалатыны жайлы мәліметтер беретін құжаттық-фактографиялық ақпараттық жүйе түрінде, ал екіншіден, арнайы мәтіндерге арналған *автоматтанған ақпараттық сөздік-анықтағыш* түрінде ұйымдастырылуы қажет.

Терминологиялық деректер банкі екі түрлі аспектіде қарастыруға болады. Ең алдымен, бұл автоматтанған ақпараттық жүйенің белгілі бір түрі деуге болады. Бірінші аспект бойынша терминологиялық деректер банкі ғылыми-техникалық лексикография тұрғысынан қарастырылатыны белгілі. Сондықтан мұндай көзқарасты ұстаным ететін болсақ, дәстүрлі терминологиялық сөздікті түзудегі мәселелермен айналысуға тура келеді. Яғни пәндік білім аясын таңдау мен оларды шектеу, ғылым тарихына қатынасы, тілдерді, материалдарды, сөзтізбені қалыптастыру әдісін және т.б. таңдап алу мәселелерімен айналысуды қажет етеді.

Автоматтандыру әдісі тек ғылыми-техникалық лексикографиядағы, әсіресе, терминологиялық сөздіктерді құрастыру ісіндегі жинақталған материалдарды өңдеу мен сұрыптау үшін ғана қажет емес, сонымен бірге сөздік мақаланы құрастыруда да аса қажет. Семантикалық жағынан қарастырғанда терминдер айтарлықтай анық топтарды құрайды деуге болады. Мысалы, оған жататындар: құрал-аспаптардың, станоктардың, тетік пен құрылғылардың және процестердің атаулары. Бұл аталғандардың өздері көптеген нақты түрлерге бөлінеді. Ал мұндай мәліметтер белгілі термин семантикасын сипаттауға қажетті өлшемдердің нақты моделін алдын ала ұсынуға мүмкіндік жасайды. Мысалы, көптеген аспап-құралдарды сипаттау үшін олардың мынадай сипаттары ескерілуі қажет: атқаратын

қызметі, тұратын орны, бөлшектер құрамы, әрекет ету принципі мен зонасы. Ал биологиялық түрді сипаттауға қажеттілер – тегі, жынысы, морфологиялық және биологиялық ерекшеліктері, ареалы (тараған ауданы), шаруашылық қызметі. Аталған параметрлер жалпы және жекелік сипатта да айырым табады. Сонымен қатар әрбір білім саласының өзіне ғана тән қасиеттері, процестері, әрекет аясы және т.б. ажыратылады. Сипаттау параметрлерін олардың терминдерді анықтаудағы ұстайтын ұстанымы бойынша және ғылымның ерекшелігіне, сөздіктің арналуына қарай топтау жүргізуге де болады.

Бірақ терминологиялық деректер банкі, сонымен бірге – құжаттық-фактографиялық типтегі автоматтанған жүйе екені айқын. Міне, осының салдарынан сөздік түзуге, құжаттарды индекстеу әдістемесін анықтауға, компьютерлік бағдарламаларды жазуға, компьютер мен тұтынушы арасындағы диалогты іске қосуға қатысты мәселелермен айналысу керек болады. Аталған проблемалық мәселелердің ойдағыдай шешім табуына байланысты өндірістік және өндірісаралық терминологиялық деректер банкі – барынша терең, семантикалық жағынан алғанда күшті құжаттық-фактографиялық сөздік-анықтағыш түріндегі диалогтық жүйе.

Болашақта түсіндірме және аударма сөздіктерді, анықтағыштарды, рубрикаторлар мен классификаторларды құрастыру жұмыстарының және терминологияны қалыпқа келтіру, терминологиялық комиссия жұмысына қызмет ету, ғылыми-техникалық аударма жұмыстарының бірден-бір негізгі бөлігі – құжаттық-фактографиялық терминологиялық деректер банкінің құрама түрі болатыны сөзсіз.

Қазіргі кездегі есептеу техникасының, қолданбалы лингвистика мен ақпараттану салаларының табыстары мен жетістіктері мұндай автоматтанған жүйелердің үнемді және тиімді болу мүмкіндіктеріне нұсқайды.

Негізінде терминологиялық деректер банкісінің жадына сөз-терминдер жайлы, соның ішінде сөз таптарына, сөз тарихына, мәтіндегі қолданысына сілтеме жасалғандарын қосып есептегенде, көптеген басқа да мәліметтерді енгізуге болады. Терминологиялық деректердің банкілер жүйесі терминдік сөзге қатысты мұндай әр деңгейдегі мәліметтерді анықтай отыра,

термин табиғатында бірге қайнасып жатқан ұғымдық мәліметті де қамтиды.

Әдебиет

Беляева Л.Н., Герд А.С., Убин И.И. Автоматизация в лексикографии. «Прикладное языкознание» учебник. Отв. ред. А.С. Герд. СПб.: Изд-во С.Петербург. университета, 1996. С. 318-333.

Лингвистическая концепция терминологического банка данных машинного фонда русского языка // Под ред. А. С. Герда. М., 1989.

Марчук Ю.Н. Проблемы машинного перевода. М., 1983. 232 с.

Марчук Ю.Н. Машинный перевод как информационная и технологическая реальность. М., 1988.

Перстнев И.П., Волкова И.Н. Банки терминологических данных. М., 1987.

Попов Э.В. Общение с ЭВМ на естественном языке. М., 1982. 360 с.

Ritchie G.D., Hanna F.K. Semantic Networks – a General Definition and a Survey // Information Technology. 1983. Vol. 2, № 4. P. 187-231.

Vauquios B. Aspects of Mechanical Translation in 1979: Conference for Japan IBM Scientific Program. Grenoble, 1979. 50 p.

Vauquios B. La traduction automatique a Grenoble // Documents de linguistique quantitative. № 24. Paris, 1975. 184 p.

Wang W.S. – L., Chan S.W., Ts'ou B.K. Chinese Linguistics and Computer // Linguistics. 1973. 118. P. 89-117.

Реферат құрастырудың әдіс-тәсілдері

Ақпарат көлемін ықшамдауға қатысты аннотациялау мен рефераттау ұғымдары. Ғылыми және ғылыми-техникалық ақпарат көлемінің үнемі ұлғаю сипатында болуына байланысты, оны тұтынушының өзіне қажетті деген мәліметтерді түпнегіз құжаттар жиынтығынан дәстүрлі іздеу тәсілімен тез арада тауып алу әрекеті қиындыққа соқтыруда. Сондықтан түпнегіз құжат бойынша екінші нұсқадағы құжаттар жиынтығын құрастыруды қажет етеді. Әдетте, оларға библиографиялық сипаттаулар, аннотациялар, рефераттар және ғылыми аудармалар материалдарын жатқызуға болады. Екінші реттегі құжат — түпнегіз құжаттағы ақпарат мазмұнын ықшамдау немесе «қысудың» (компрессиялаудың) нәтижесі болып табылады. Мұндай қысу мен ықшамдау библиографиялық сипаттаулар үшін өте-мөте тиімді, себебі мәтін орнында тек олардың тақырыптарының аттары ғана қалып отырады. Ал ғылыми аудармалардың көлемін «қысудан» ұтатынымыз аз. Себебі, аударма құжаттағы ақпараттың көлемі мен оның мазмұны түпнұсқа құжатпен барабар болуы қажет. Соңғы кезде кейбір аудармалардағы айтылымдарды қысқарту арқылы көлемді ықшамдау әдісі тәжірибеге енуде. Бірақ мұндай әдіспен де көлемді ықшамдау тұрғысынан ұтатынымыз аз. Ақпарат мазмұнынан аннотация мен реферат алу кезіндегі мәтін көлемінің «қысылу» нәтижесі жоғарыда аталған библиография мен ғылыми аударма

құжаттарымен салыстырғанда көлемдік жағынан олардың орта шенінде деуге болады.

Аннотация мен реферат бір-бірінен ақпарат мазмұнынан айырым табады. Аннотацияның қызметі — толығымен нұсқағыштық. Себебі, аннотацияда түпнегіз құжаттағы ақпарат мазмұнын толық ашпаса да, сол мазмұнға нұсқау тәсілі орын алады. Ал реферат қызметіне келетін болсақ, ол ақпараттың мазмұнына нұсқаумен бірге, қысқаша ақпарат беру рөлін де атқарады. Яғни түпнегіз құжаттың мазмұны рефератта (азды-көпті болса да) толық дәрежеде болуға әрекет етеді. Ағылшын тілді елдерде аннотация мен реферат (indicative abstract) құрастыруда ақпараттың көлемдік өлшемі алынады. Аннотацияның таза нұсқағыштық қызметі аз көлемді қажет етсе, ал реферат оған қарағанда (өз рөлін толық орындау үшін) әлдеқайда көп көлемді қажет етеді. Әсіресе мұндай жағдай гуманитарлық ғылымдарға қатысты рефераттарға тән. Алайда рефераттардың көлемін дәл анықтау мүмкін бола бермейді, себебі ол көптеген факторларға байланысты болып келеді. Мысалы, оған жататындар түпнегіз құжаттың сипаты (мақала, монография және т.б.), түсініктілігі (түсінуге қиын құжаттар толығырақ рефератталады), түпнұсқа құжаттың тілі (сирек пайдаланылатын тілдердегі құжаттар егжей-тегжейлі рефератталады), ғылыми типі немесе пәндік ая түрі (жаратылыстану ғылымы немесе қоғамдық ғылымдар аясы) және т.б.

Интеллектуалды рефераттау. Түпнұсқа құжат мазмұнын интеллектуалды рефераттау ұғымына адам баласының ой қызметінің әрекеті арқылы, әдетте, маман адам (референт) арқылы орындалатын рефераттау әрекетін атайды. Қазіргі кезде рефераттық журналдардың басым көпшілігінде осындай әдіс орын алады.

«Интеллектуалды» терминін бейнелейтін ұғымның негізінде рефераттау жұмысын жүзеге асыратын субъектінің адам екені мәлім. Сондықтан интеллектуалды рефераттау әдісіне ЭЕМ көмегімен атқарылатын автоматты рефераттау әдісі қарсы мағынада айтылып жүр және соңғы әдістің, яғни автоматты рефераттау әдісінің рөлі мен сұранысы барынша артуда.

Рефераттау мәселесін жеке қарастырғанда автордың өзі жазған рефераты (summary) мен сол саланың маман референтінің рефератын (abstract) ажырата білу аса маңызды. Бірінші рефератты түпнегіз құжат бойынша автордың өзі құрастырады және онда құжаттың қалам иесінің түсінігіндегі ең маңызды деген аспектілері бейнеленеді. Әдетте, авторлық реферат түпнұсқа құжатқа (мысалы, мақалаға немесе монографияға) қысқаша шолу ретінде көрініс табады. Бірақ мұндай реферат әр уақытта жасала бермейді және тіпті де барлық авторларға бірдей тән емес.

Авторлық рефераттың қызметі – мәтіннің мазмұны жайлы оқырманды ақпараттау болып табылады. Мұндай ақпараттау нәтижесі аннотация арқылы берілетін нұсқағыш ақпараттан аз айырым табады. Авторлық реферат ешбір іздестіру қызметін атқармайды. Ал әдеттегі рефераттық журналдардағы рефераттар іздестіру қызметін де атқарады. Себебі «қолмен іздеу» қызметі кезінде құжаттың релеванттық критерийі (мәндік өлшемі) рефераттың мазмұнына қатысты анықталады. Кейбір өндірісі дамыған капиталистік мемлекеттердің рефераттық журналдарында авторлық рефераттар көбінде ашық түрде жарнамалық сипатқа ие. Сондықтан авторлық рефераттар бойынша құжаттың негізгі мазмұны жайлы пікір айту қиынға түседі.

Авторлық рефераттар көлемі жағынан, әдетте, референттердің құрастырған рефераттары көлемінен айтарлықтай аз болады. Көпшілік жағдайларда мұндай рефераттарда зерттеу жұмыстарының нақты нәтижелері келтіріле бермейді.

Рефераттық журналдар (РЖ). Қайсыбір пәндік аяны немесе келелі мәселені бейнелейтін библиографиялық сипаттаулар олардың тақырыбы бойынша немесе автордың аты-жөні бойынша реттелуі мүмкін. Әдетте, мұндай әрекет *әдебиеттердің библиографиялық көрсеткіші* деп аталатын кітап түрінде дайындалады. Библиографиялық сипаттаулардың бәрінде бірдей (немесе тек кейбіреулерінде) аннотация берілуі мүмкін. Осының нәтижесінде әдебиеттердің аннотацияланған библиографиялық көрсеткіші құрастырылады.

Әрбір ғылым саласына қатысты библиографиялық сипаттаулар каталогтық кәртiшкілер түрінде де көрініс табуына жағдай жасалады. Каталогтық жәшіктерге бұл кәртiшкілер

автордың аты-жөні бойынша немесе білім салалары арқылы әліпби тәртібіменен реттеліп қойылса, олар *авторлық* немесе *жүйелік каталогтар* деп аталады. Бұларды әдебиеттердің библиографиялық көрсеткішімен бірдей дәрежедегі қолмен орындалған (автоматтанбаған) *ақпарат-іздестіру жүйелері* деп те қарастыруға болады.

Рефераттық журналдардағы рефераттар да осы айтылғандарға ұқсас түрде жүйеленеді. Бұлайша жүйелеудің негізінде – білім универсумын (әлемдік білімді) қалаған дәлдік дәрежесін сақтай отыра, салалар мен пәндік аяларға бөлу принципі жатыр. Қажетті деген мәліметті тауып алу әрекетін барынша жеңілдететіндіктен мұндай реферативтік журналдар да қолмен жүзеге асатын (автоматтанбаған) *қарапайым ақпарат-іздестіру жүйесінің* қатарына жатады.

Қазіргі кезде, әлемдік деңгейде қарастырғанда, тақырыбы мен құрылымы жағынан өзгешеленетін көптеген реферативтік журналдар қолданыс тауып жүр. Мысалы, Кеңес Одағы кезіндегі реферативтік журналдардың негізгілері ретінде РЖ ВИНТИ және РЖ ИНИОН түрлерін атауға болар еді Әрине, одан басқа да реферативтік журналдарды, мысалы, РЖ ВНИИМИ, РЖ ВИНТИСХ және т.б. осылардың қатарына қосуға болады.

РЖ ВИНТИ жаратылыстану мен дәл ғылымдар құжаттарын, РЖ ИНИОН гуманитарлық ғылымдар салаларын бейнелейді. Реферативтік журналдардың деректері сыртқы пішін жағынан – форматы бойынша, ал іштей – рефераттардың құрылымы мен көлемі жағынан айырым табады. РЖ ВИНТИ рефераттары РЖ ИНИОН-ға қарағанда көлем жағынан барынша шағын келеді және канонды (тұрақты) қалыпқа ие болып келеді. Басқаша айтқанда, рефераттың бастапқы жағында түпнұсқа құжатында сипатталған мәселелерге нұсқау жасалады да, ал қорытынды бөлігінде алынған нәтижелер аталады. РЖ ИНИОН көлемі РЖ ВИНТИ рефератының көлеміне қарағанда бірнеше есе көп (10 есе және одан да көп) және құрылымы жағынан шағын мақаладан ешбір айырмашалығы болмайды. Себебі, қоғамдық ғылымдар мәселесін бірнеше тармақтарға қисындастыру немесе шағын түрде қорытындылау мүмкін бола бермейді, ал ғылыми тұрғыда дәлелділік сипатының орын алуы аса маңызды деп саналады.

Рефераттық журналдарда кез келген құжаттың рефераты жеке-дара түрде берілмейді. Ол автордың аты-жөнін, бас тақырыпты және шығу деректерін қоса есептегендегі библиографиялық сипаттауды да бірге қамтиды. Мұндай жағдай библиографиялық сипаттау мәтіні мен реферат мәтінін әртүрлі тұрғыда реттеуге мүмкіндік тудырады. Рефераттардың ең көп тараған үш түрін атауға болады:

1) мынадай реттегі (ағылш. conventional abstract): бас тақырып – библиографиялық сипаттау – реферат мәтіні түріндегі әдеттегі реферат;

2) бірінші орында тақырыпқа қатысты сөйлем орналасқан реферат. Ақпараттың реті мынандай: тақырыптық сөйлем – реферат мәтіні – библиографиялық сипаттау. Мұндағы «тақырыптық сөйлем» дегенді, ақпараттық тұрғыдан қарастырғанда, ең маңызды деп саналатын реферат мәтінінен алынған сөйлем деп ұғыну керек;

3) бірінші орында «бас тақырып» орналасқан реферат. Оның реттілігі мынадай: бас тақырып – тақырыптық сөйлем – реферат мәтіні – библиографиялық сипаттау (бас тақырып қайталанбайды).

Әртүрлі реферативтік журналдарда осы аталған рефераттардың қайсыбір түрі қолданылады. Мысалы, «Chemical Abstract» реферативтік журналда рефераттың 2-типі, ал ВИНТИ атты реферативтік журналда – рефераттың 1-типі қолданылады.

Рефераттау жайындағы әдебиеттерде интеллектуалды рефераттың ең ұтымды (оптималды) нұсқасын тандауға қатысты мәселе бірнеше рет талқыланды. Оны реферат ішінде түпнегіз құжат мазмұнының қай құрамдас бөліктерінің қандай реттілікте көрініс табуының қажеттігі жайлы мәселе деуге болады. Бірақ көптеген тәжірибелердің нәтижелеріне сүйенсек, алдын ала берілген қайсыбір түпнегіз құжат рефератының ұтымды нұсқасы болмайды. Шындығында, түпнұсқа құжаттың бір ғана мәтіні бойынша әртүрлі рефераттар құрастыруға болады. Және мамандар тұрғысынан олардың бәріне де қанағаттанарлықтай деген баға берілуі мүмкін. Сондықтан аталған мәселеге соңғы кезде басқаша түрде түсініктеме беріліп жүр: рефераттың ең ұтымды нұсқасын іздестіруден гөрі, құжаттың бір ғана мәтіні

бойынша белгілі бағыттағы мамандар қажетін қанағаттандыра алатын бірнеше реферат құрастырылғаны жөн. Құжат мәтінінен реферат алу мәселесінде осындай көзқарасты ұстаным ету автоматты рефераттау жүйелерін құрастыруда көптеп қолданыс таба бастады.

Автоматты рефераттау. Автоматты рефераттау жүйесін құрастырумен айналысатын көпшілік ғалымдар түпнегіз құжаттың мәтінінде ақпараттау фрагменттерінің қайсыбір түрі орын алады деген тезиске сүйенеді. Сондықтан автоматты рефераттаудың негізін құрайтын ең ақпаратты деген фрагменттерді айқындау зерттеу жұмысының басты міндеттерінің бірі болып табылады.

Қазіргі кездегі түсінікке сәйкес, ақпараттың көпшілігі мәтіннің мынадай тұстарында шоғырланады:

- 1) мәтіннің бір нүктесінде;
- 2) мәтіннің бірнеше фокалды немесе тірек нүктелерінде;
- 3) мәтіннің кез келген жерінде кездесуі мүмкін, сондықтан

оның қатаң түрде анықталған орны жоқ.

Ең көп тараған көзқарас бойынша, өзінде барынша көп мөлшердегі ақпаратты сақтайтын мәтін орны – тірек нүктелер. Ондай нүктелер, әдетте, мәтіннің басында немесе соңында орналасады. Дәлірек айтсақ, ондай орындарға: тақырып, кіріспе, тұжырымдар, қорытынды, ұсыныстар жатады. Ал аталғандардың ішінен ең негізгісі деп тұжырымдар мен қорытындыны атауға болады. Бірақ бұдан да басқа пікір кездеседі. Мысалы, ол бойынша ең көп ақпарат бөліктері мәтіннің барлық жерінде азды-көпті болса да бірдей дәрежеде үлестіріледі.

Қандай көзқарасты ұстаным етсек те мәтіннің ақпараттық бөліктерін айқындау үшін барынша мұқият зерттеуден өткен әдіс-тәсілдер жиынтығы болуы керек. Қазіргі кезде мәтін мен оның бөліктерінің әртүрлі белгілеріне сүйенетін көптеген әдіс-тәсілдер қолданылып жүр. Бұлардың бәрі бірдей мәтінді үзінді бөліктерге бөлшектеуді ұстаным етеді: ең алдымен мәтін кесінділерге бөліктеледі, сосын олардың ішінен ең ақпаратты деген сөйлемдер, сөз тіркестері және сөздер іздестіріледі. Осындай әдіс-тәсілдердің кең қолданыс тауып жүрген түрлеріне жеке тоқталайық.

Статистикалық-дистрибутивті әдістер. Статистикалық-дистрибутивті әдістерді басқаша түрде – тілдік бірліктердің (морфема мен сөздердің) мәтін ішіндегі үлестірілу заңдылығын (тізбектеліп келу заңдылығын) статистикалық зерттеу әдісі деп те атайды. Аталған әдіс мына жайтқа негізделеді. Қарастырылып отырған тілдік бірліктің мүмкін болатын барлық қоршаулар жиынтығы мен олардың сандық деректерін (немесе статистикасын) дистрибуция (тіркесімдік) тұтастығы деп қарастыру. Ал тәжірибе арқылы алынатын статистикалық өлшем статистикалық-дистрибутивті әрекеттердің объективтілігі мен алгоритмделуін, яғни бұл әдіс-тәсілдердің автоматтану мүмкіндігін қамтамасыз етеді.

Статистикалық-дистрибутивтік әдіс-тәсілдер дешифровка жасауда (құпия жазуды тануда), статистикалық-комбинаторлық модельдеуде, автоматты индекстеу мен рефераттауда және семантиканы зерттеуде кең қолданыс тауып отыр.

Автоматты жолмен рефераттау мақсатында қолданылатын жалпы статистикалық-дистрибутивтік әдіс-тәсілдер ішінен бірнеше жеке әдістерді бөліп қарастыруға болады. Ең алдымен ондай жеке әдіске ғалым Х.П.Лунның ұсынған әдісін жатқызуға болады. Ол бойынша қарастырып отырған құжат үшін ең ақпаратты деген сөйлемдер қатарына бір-біріне барынша тығыз орналасқан мәнді сөздер жиынтығын қамтитын сөйлемдер жатады. Бұл жағдайда сөйлемдердің мәтіндегі орналасу орны ескерілмейді, сондықтан ақпаратты сөйлемдер түпнегіз құжаттың кез келген жерінен алынуы мүмкін. Сөйлем ішіндегі сөздердің шоғырлануы (топтануы) деп, қарастырып отырған мәтін үшін шеткі сөздері мәнді болып келетін кез келген сөзтұлға тізбегін атайды. Х.П.Лун әдісі бойынша, сөздер арасындағы мағыналық қатынас ескерілмесе де, ол толық автоматтандыру мүмкіндігімен ұтымды деп саналады.

Лун әдісін оңтайландыру мақсатында, оның бірнеше түрленімдері жасалған. Мысалы, В.А.Освальдтың әдісі бойынша, сөйлемнің ақпараттылық деңгейінде тек сөздердің мәнділік дәрежесіне қарай шоғырлану мүмкіндігі ескеріледі. Басқаша айтқанда, бұл әдісте тірек сөздердің топталуы ғана ескеріліп қоймайды, сонымен бірге осындай шоғырланудың сандық сипаты да есепке алынуы керек болады. Ал Л.Эрлдің

әдісі бойынша, сөйлемдердегі сөздердің шоғырлану саны үшке тең немесе үштен көп болса ғана ондай сөйлемді ақпаратты сөйлем қатарына жатқызуға болады. ACSI-Matic әдісінде шоғырланған тірек сөздер жиілігі құжат ішіндегі сөздердің орта жиілігінен үлкен болатын шамасы ғана ескеріледі. Л.Дойл мен М.Квиллиан мынадай әдістемені ұсынды: сөйлемдерде терминдердің немесе тірек сөздердің бірге кездесу мүмкіндігін есепке алу мен арнайы матрица арқылы өзара байланыста болатын элементтердің ұқсастық коэффициентін есептеу.

Мұндай «бірге кездесу» принципін ұстаным ету басқа да ғалымдардың (мысалы, А.Я.Шайкевичтің) зерттеулерінде кездеседі. Бірақ ол ғалымдар бұл әдіс-тәсілді рефераттауды автоматтандыру мақсатында емес, тек семантиканы зерттеу әдісі ретінде қолданғаны белгілі.

Статистикалық-дистрибутивтік әдіс мәтін ішінен ең ақпаратты деп саналатын сөйлемдерді қарапайым жолмен бөліп алып, оларды рефератқа енгізуге мүмкіндік тудырады. Мұндай сөйлемдерді іздестіру аясын шектеу мақсатында статистикалық-дистрибутивтік әдіс позициялық (орналасу тәртібін ескеретін) әдіспен бірге қолданылуы да мүмкін.

Статистикалық-дистрибутивті әдістің кемшілігі ретінде мәтін үзіндісінің ақпараттылығы мен олардың құрамындағы тірек сөздердің арасында тікелей және бірімәнді байланыстың болмауын айтуға болады. Әрине, мұндай байланысты анықтау ісі барынша күрделі мәселенің қатарына жатады. Сондықтан бұл әдіс, бір қарағанда, аса қарапайым болып көрінгенімен, зерттеушіні қалаған нәтижесіне жеткізе бермейді. Сонымен бірге тек қана статистика-дистрибутивті әдістемеге негізделіп құрастырылған реферат әр кезде бірдей дәрежеде жеткілікті түрде қисынды сипатта көрініс таба бермейді. Аталған рефераттың интеллектуалды рефератқа қарағанда кемшілік жақтары жеткілікті. Сондықтан қажеттілікке қарай, статистикалық критерий басқа критерийлермен толықтырылып отыруы керек.

Мағыналық аспектіні қолдану әдісі. Статистикалық-дистрибутивті әдіс бөліп алынған мәтін үзіндісінің мазмұндық жағы туралы ешбір мәлімет бермейді. Сондықтан оның ақпараттық сипаты жайлы тек жобалап қана айтуға болады.

Бірақ бөлікке бөлініп алынған мәтін үзінділерінің мән-мағыналық жағын арнайы жүргізілетін әрекеттер арқылы нақтылауға болады. Мұндай әрекет түріне мәтін үзіндісін мағыналық индикаторлар арқылы белгілеу барысын жатқызуға болады. Аталған индикаторларды көпшілік жағдайда мағыналық аспект деп атайды да, ал оған негізделетін рефераттау әдісін аспекттік әдіс дейді.

Әр автор әртүрлі аспектілер жиынын ұсынуы мүмкін, бірақ солардың ішінен жиі қолданылатындары: зерттеу пәні, мақсаты мен зерттеу әдісі, актуалдық, ұсынылған әдістің ерекшеленуі немесе оның шешімдері, артықшылығы, қолдану аясы, түйіндері, нәтижелері, ұсыныстар және т.б. Кейбір жағдайларда мұндай аспектілер семантикалық топтарға біріктіріледі.

Рефераттаудың аспекттік әдісінің құндылығы белгілі. Ол әдіс бойынша рефератқа белгілі мазмұндағы және ең ақпаратты деп қарастыруға болатын сөйлемдер мен мәтін үзінділерін енгізуге мүмкіндік туындайды. Көптеген зерттеушілердің бақылауынша, ең ақпаратты деп саналатын аспектілерге «шешімге жетудің ұсынылған вариантын» және «түйіндерін» жатқызуға болады. Маңызды аспектілер қатарына, көбінде, тақырыпта көрініс табатын мән-мағыналар жатады.

Аспекттік әдіс түпнегіз құжат мәтінінен мәнділік зоналарды бөліп алуға және олардың әрбіреуіне салмақтық коэффициент шамасын тіркеп жазуға мүмкіндік туғызады. Салмақтық коэффициенттің шамасына қарай, сол зона үзіндісін рефератқа енгізу не енгізбеу мәселесі шешіледі. Аспекттік әдісті статистикалық-дистрибутивті әдіспен бірге де қолдануға болады.

Аспекттік әдіс бойынша құрастырылған рефераттар бір-бірімен айтарлықтай байланыста болмаса да, тұтынушыға жеткілікті дәрежеде ақпарат беретін бірнеше мәтін үзінділерінен тұрады. Мәтін үзінділерінің байланыссыздық дәрежесін *мәтіндік байланыстар әдісін* қолдану арқылы төмендетуге болады.

Мәтіндік байланыстар әдістері. Мәтіндегі сөйлемдердің ақпараттық мәнділігі тек олардың бөліктерінің статистикалық сипаттары мен мазмұндық аспектілері арқылы ғана анықталып қоймайды, ол сонымен бірге басқа да сөйлемдердің байланыстық сипаты мен санына қатысты болады. Көптеген ғалымдар-

дың (В.Е.Берзон, С.М.Приходько, Э.Ф.Скороходько, И.П.Севбо және т.б.) ұсынған автоматты рефераттау моделі осындай ұстанымда құрылған.

Енді ақпараты мол және мәнділігі жоғары деп саналатын сөйлемдерді тандау әдісін қарастырайық. Бұл әдіс бойынша байланыстардың санын білдіретін қайсыбір шекаралық (пороговая) шама алдын ала берілуі қажет. Осы шаманы реттеу арқылы сөйлемдердің ақпараттық қанықтығын айқындауға мүмкіндік бар. Мұндай әрекет бойынша рефераттың көлемін өзгертуге болады. Әрине, мұндай жағдайда сөйлемдердің сипаты ескерілмеуі де мүмкін.

Бұл әдістің тағы бір жақсы нұсқасында мәтіндегі фразадан тыс бірліктердің (сверхфразовые единицы) *эксплицитті* және *имплицитті* түрлері өзара ажыратылып қарастырылады. Мұндай жағдайдан, әдетте, негізгі мазмұн фразадан тыс бірлікке *эксплицитті* болып келетін бірінші сөйлемнің үлесіне қатысы бар. Бұл сөйлем, көпшілік жағдайда, автосемантикалық болып келеді. Жоғарыда сөз болған *фразадан тыс бірлікке эксплицитті* келетін «бірінші сөйлемнің» ақпараттылығы сол бірліктің ұзындығының артуына қарай ұлғайып отырады.

Фразалық байланыс өлшемдерін ескеру ісі реферат нәтижесін барынша *байланыстық* және *толықтық* сипатта бейнелейді. Бірақ осындай жағдайдың өзінде де аталған реферат сапалық жағынан интеллектуалды рефераттан көп төмен тұрады.

Құрамдастырылған әдістер. Мәтіннің ең ақпаратты деген үзіндісін бөліп алуға байланысты қолданылатын әрбір жеке әдістің іштей өзіне ғана тән шектеулері болатыны белгілі. Рефераттау әрекеті кезінде тұтынушыны қанағаттандыратындай нәтижеге жету үшін әр әдістің тиімді жақтарын бөліп алып, оларға құрамдастыру әдісін қолданған жөн болады. Мәселен, зерттеушілердің барлығы бірдей мәтін ішіндегі сөздің немесе оның үзіндісінің кездесу жиілігін ескерудің маңыздылығына нұсқайды. Сол сияқты мәтін бойындағы сөйлем бірлігін немесе оның басқа бір үзіндісін, ал басқа бір жағдайда олардың мәтіндік байланыстылығын тірек етіп алу ұстанымының маңыздылығы да алдыңғы әдістен ешбір кем емес. Тағы бір ғалымдар тірек сөздердің аралық қашықтығын, тақырып

құрамына енген тірек сөздерді және мәтін типін ескеру қажеттігін де баса айтады. Ал морфологиялық тұрпат (форма) пен сөздің немесе сөз тізбектерінің синтаксистік қызметі жайлы факторлар өте сирек сөз болады.

Нәтижелік рефератқа енуге тиісті ақпараттық мәнді мәтін бірліктерін немесе үзінділерін бөліп алу әдістері белгілі реттілікпен қолданылуы қажет. Мәселен, лексиканың статистикасын анықтауға арналған әдістер ең ақпаратты деген сөйлемдерді немесе олардың жиынтығын мәтін ішінен бөліп алу үшін қажет. Сонымен бірге мұндай бөліп алу барысы шептік (позициялық) және аспектілік әдістер арқылы нақтыланып, түзетіліп отырулары қажет. Бірақ рефераттың мәтіндік қисындылығына фразеаралық байланыс өлшемін қолдану арқылы ғана қол жеткізуге болады. Сондықтан рефераттау әдістерін қолдану тәртібі төменнен жоғарыға қарай бағытталған табиғи қозғалыс сызбасы бойынша орындалады. Ондай сызба мынадай реттілікке негізделеді: лексика – сөйлем – шебі (позициясы) немесе аспектілі келісілген сөйлем жиындары – фразеаралық байланыстар – реферат. Бірақ бұл сызбаға қарама-қарсы бағытпен де, яғни жоғарыдан төменге қарай бағытталған сызба түрі де қолданылуы мүмкін. Ол бойынша алдымен түпнегіз құжаттың толық мәтініне бастапқы үзінділерге бөлшектеу тәсілі жүзеге асырылады да, сосын барып үзінділер аясында орныққан орнына қарай немесе аспектілік жағынан сипатталған сөйлемдер бөлініп алынуы мүмкін. Ең соңында лексикалық деңгейге өту жолы қарастырылады. Аталған екі әдісті бір-біріне қарама-қарсы бағыттағы екі әдіс-тәсіл ретінде жарыстыра қолдануға да болады.

Интеллектуалды және автоматтау әрекеті бойынша құрастырылған рефераттар. Интеллектуалды рефератты құрастырудың мақсаты – тұтынушыны қажетті ақпаратпен қамтамасыз ету немесе оның іздестіру қажеттілігін қанағаттандыру. Басқаша айтқанда, рефераттарды пайдаланатын адам болғандықтан, яғни оның ақпараттық қажеттігін өтеу. Әрине, мұндай рефераттар басқа да мақсаттар үшін пайдаға асуы мүмкін, мысалы, құжаттарды қол күшімен немесе автоматты түрде индекстеуге, терминологиялық сөздіктер мен ақпарат-іздестіру тезаурустарын құрастыруға, мәтіндердің атрибуциясын

(шығарма авторы, шыққан мезгілі мен орны) анықтау және т.б. мақсаттар үшін қолданыс табуы мүмкін. Бірақ бұл аталған мақсаттардың барлығы да интеллектуалды рефераттар үшін негізгі емес, жанама деректер.

Ал автоматты түрде рефераттаудың мақсаттары интеллектуалды рефераттау мақсатынан көпшілік жағдайда ерекшеленеді. Әрине, мұндай рефераттар да тұтынушының, яғни адамның мұқтажығын қанағаттандыру үшін арналған болуы мүмкін. Бірақ машина арқылы жүзеге асатын рефераттардың мүмкіншілігі әзірше интеллектуалды рефераттардың мүмкіншіліктерімен салыстыруға келмейді.

Машиналық рефераттар көбінде индекстеу мақсатында қолданылады. Сонымен бірге мұндай мақсатқа байланыссыз немесе байланысы аз сөйлемдердің немесе олардың үзінділерінің тізбегі түрінде көрініс табатын рефераттар да сай келе беруі мүмкін. Аталған рефератқа қойылатын негізгі талап – құжатты индекстеуге қажетті тірек сөздер санының жеткілікті болуы. Мұндай рефераттар, әдетте, квазиреферат деген терминмен аталып жүр. Квазирефераттардың қолданылатын екінші бір саласы ретінде мәтіннің атрибутын атауға болады. Бұл бағыт бойынша мәтінге пәндік салаға сәйкес келетін индекстер тіркеліп жазылуы қажет. Бұл жағдайда рефераттың байланыстылық сипатының жоғары дәрежелі болуы қажет те емес.

Сонымен, квазирефераттар дегеніміз өзінше бір лексикалық тұрғыда өзекті саналатын рефераттар, ал интеллектуалды рефераттар – дискурс бойынша өзекті рефераттың түрі. Басқаша айтқанда, олар байланысты мәтін өлшемдерін қанағаттандыра алатын реферат түрі деуге болады.

Бұл аталған айырмашылықтар аса маңызды болғанымен, жоғарыда айтылғандардан машиналық рефераттың өлшемдерін интеллектуалды рефераттың параметрлеріне жуықтататын автоматтанған амалдар жүйесін құру мүмкін емес деген қорытынды шықпайды. Мұндай мақсатқа жету үшін автоматты түрде рефераттаудың әртүрлі әдіс-тәсілдер мен амалдар жиынтығын қолданумен бірге, олардың үйлесімділік жағын да ескеру қажет болады. Қорыта айтқанда, мұндай бағытты ұстаным ету – әрбір тұтынушының талабына сай келеді және қажетті деген прагматикалық бағытты қамтамасыз ете алады.

Э д е б и е т

Бондаренко Г.В. Об использовании формальных методов анализа структуры целого текста для автоматического индексирования и реферирования // Научно-техническая информация. Сер. 2. 1975. №5.

Гендина Н.И. Выявление аспектов содержания документов с помощью формальных тестовых признаков // Научно-техническая информация. Сер. 2. 1979. №4.

Гиндин С. И. Методы автоматического фрагментирования текста, опирающиеся на характеристики внутреннего состава фрагмента // Семиотика и информатика. Вып. 9. М., 1977; Позиционные методы автоматического фрагментирования текста, их теоретико-текстовые и психолингвистические предпосылки // Семиотика и информатика. Вып. 10. М., 1978.

Горькова В.И., Борохов Э.А. Реферат в системе научной коммуникации: Направления совершенствования лингвистических и структурных характеристик // Итоги науки и техники. Сер. Информатика. М., 1987.

Добрускина Э.М., Берзон В.Е. Синтаксические сверхфразовые связи и их инженерно-лингвистическое моделирование. Кишинев, 1986.

Карасев С.А. Построение системы автоматического реферирования, включающей анализ и синтез текста (АНАСИН) // Научно-техническая информация. Сер. 2. 1978. №5; Система автоматического реферирования: результаты машинного эксперимента // Научно-техническая информация. Сер. 2. 1978. №8.

Леонов В.П. О методах автоматического реферирования (США, 1958-1974) // Научно-техническая информация. Сер. 2. 1975. №6.

Новиков А.И., Якушин Б.В. Алгоритм индексирования текстов взвешенными ключевыми словами по методу семантической фильтрации // Научно-техническая информация. Сер. 2. 1972. №6. С. 15-20.

Пащенко Н.А., Кнорина Л.В., Молчанова Т.В. и др. Проблемы автоматизации индексирования и реферирования // Итоги науки и техники. Т.7. М., 1983.

Севбо И.П. Структура связного текста и автоматизация реферирования. М., 1969.

Скороходько Э.Ф. Семантические связи в лексике и текстах // Вопросы информационной теории и практики (М), 1974. №23.

Шайкевич А.Я. Дистрибутивно-статистический анализ в семантике // Проблемы и методы семантических исследований. М., 1976.

Ақпарат іздестіру жүйесі

Соңғы 20 жыл ішінде танымдық ой-сана мен өндіргіш күштердің дамуы нәтижесінде өмірге жаңа ғылым саласы — информатика ілімі келді.

Бір жағынан алғанда, информатика — ақпаратты ұйымдастыру мен өндеудің жалпылама заңдылықтарын қарастыратын ғылым. Екіншіден, оны бүгінгі таңдағы қоғамдық өндірістің жаңа саласы деуге де болады. Информатика ілімінің өмірге келуі мен даму деңгейлері ең алдымен ақпаратты өңдеу құралдарының, оның ішінде есептеу техникасы мен байланыс құралдарының дамуына тікелей қатысы бар. Ғылыми-техникалық прогресті жаңаша түрде ақпаратпен жабдықтандыру мәселесі автоматтанған жүйелердің кең түрде қолданыс табуымен сипатталады. Қазіргі кезде деректерді электрондық әдіспен өңдеуге қатысты жүйелер әртүрлі мәселелерді шеше алады: есептеу, технологиялық тіркеу, статистикалық тіркеу, баспа жұмыстарының есеп-қисаптары және таза ақпараттық (ақпараттық сақтау мен іздестіру) мәселелер.

Ғылыми-техникалық прогрестен туындайтын ең бір маңызды салдар — ақпарат көлемінің бұрын-соңды болмаған дәрежеде артуы. Ғылыми-техникалық басылымдар мен жарияланымдардың сан жағынан өсуі қажетті деген мәліметтерді тауып алу мәселесін кенеттен қиындатып жіберді. Осыған байланысты ақпарат-іздестіру теориясы дамып, ақпарат іздестіру жүйелері мен құрылғыларды құрастыру жұмысында жүзеге

асты. Ақпарат іздестіру жүйелерінің қол күші арқылы іске асатын механикаландырылған және автоматтандырылған түрлері болады. Ең алғашқы ақпарат іздестіру құрылғылары сұратуға (запрос) сай қажетті деген құжатты тауып алуды механикалық сәйкестендіру арқылы іске асыратын техникалық құрал түрінде жасалды. Оларда перфокарталар, суперпозициялық карталар, шеттері перфорацияланған перфокарталар және т.б. қолданыс тапты. Ақпарат іздестіру қызметінде қолданылатын есептеу техникасының дамуы *автоматтанған ақпараттық-іздестіру жүйесін* өмірге әкелді. Телекоммуникациялық құралдардың дамуы үлестірілген ақпараттық іздестіру жүйелерін, әлемдік ақпарат іздестіру жүйелерін және әлемдік ақпараттық инфрақұрылым тұтастығын жасауға өз әсерін тигізді. Біздің баяндауымызда тек формалды түрдегі сұратулар бойынша іске асатын, қажетті мәліметті бойында сақтаған мәтінді (құжат, оның бөліктері, фактографиялық жазба) іздестіруге арналған автоматтанған ақпараттық жүйелер тобы қарастырылады. Аталған жүйенің басқа жүйелерден негізгі айырмашылығы мынада. Автоматтанған ақпарат іздестіру жүйесі әрекетінің нәтижесінде алынған деректердің мағыналық мәнін ашу мен ол бойынша тиісті шешім қабылдау жақтары қарастырылмайды. Яғни ақпарат іздестіру жүйесі – мәтіндерді (құжаттарды) немесе мәліметтерді (фактыларды) сақтау мен іздестіруге арналған құралдар мен әдістердің жиынтығы болып табылады.

Ақпарат іздестірудің мәні және негізгі терминдер. Ақпарат іздестірудің мәні – *құжаттың мазмұны мен сұрату бейнесін* өзара теңдестіру. Ал екі түрлі мәтіннің мазмұнын салыстыру дегеніміз мәтіндердің мағыналық жақындық (ұқсастық) дәрежесін анықтау. Ақпарат іздестірудің негізгі ұғымы (құжат пен сұрату ұғымдарымен бірге) релеванттылық ұғымы (ағыл. relevant – *қолайлы, қатысы бар*) болып табылады. Әдетте, релеванттылық ұғымын бір мәтіннің екінші мәтінге сәйкестік дәрежесі – құжат пен сұратудың сәйкестігі деуге болады. Ал релеванттылық – мағыналық релеванттылық пен сұратудың сәйкестігі ретінде де түсінік табады. Релеванттылық сонымен бірге мағыналық релеванттылық және формалды релеванттылық түрінде де екіге ажыратылады. *Мағыналық реле-*

ванттылық — ақпаратты сұратуға байланысты табылған құжаттың ақиқаттық сәйкестігі. Ал *формалды релеванттылық* — құжаттың іздестіру бейнесінің іздестіру ұйғарымына сәйкестігі. Бұл жердегі *іздестіру бейнесі* мен *іздестіру ұйғарымының* түсінігі дегенді құжаттың негізі мазмұны мен сұратуды арнайы ақпаратты іздестіру тілінде *бейнелеу* деп ұғынған жөн. Жоғарыда аталған мағыналық және формалды релеванттылықпен бірге «пертиненттілік» (ағыл. *pertinent* — қолайлы, іске қатысты) түрін де ажыратады, яғни құжат мазмұнның шындығындағы ақпараттық мұқтаждыққа сай келуі. Ал ақпараттық мұқтаждық — белгілі индивидтің немесе ұйымның нақты мақсатқа жету жолындағы білімге мұқтаждығы. Ақпараттық мұқтаждық табиғи тілдегі мәтін түрінде көрініс табады және ол *ақпараттық сұрату* деп аталады. Ақпараттық сұрату ақпараттық мұқтаждықтың көрінісін әрқашан бірдей дәрежеде дәлме-дәл сипаттай алмайды. Себебі, ақпараттық мұқтаждықты тұжырымды етіп айтудың субъективтік сипатта болуы әртүрлі қиындықтарға әкеледі.

Релеванттылық ұғымына ақпаратты іздестірудің негізгі нәтижелік көрсеткіші, яғни толықтық пен дәлдік сипаттары негізделеді. *Іздестірудің толықтық өлшемі* деп ақпараттық жиымда (массив) қамтылған сұранысқа сай келетін релеванттық құжаттар санының құжаттардың жалпы санына қатынасын атайды. Ал *іздестірудің дәлдік өлшемі* — берілген релеванттық құжаттар санының тауып алынған барлық құжаттар санына қатынасы.

Іздестірудің нысаны мен алға қойған мақсатына байланысты екі түрлі ақпарат іздестіру орын алады: құжаттық және фактографиялық. Осыған сәйкес ақпарат іздестіру жүйесінің де құжаттық және фактографиялық түрлері айырым табады. Құжаттық іздестіруде тұтынушыға ұсынылатын нысандар ретінде қажетті мәліметтерді өзінде сақтайтын құжаттың өзі немесе олардың библиографиялық баяндалуы (описание) жайлы мәлімет берілді. Ал *фактографиялық іздестіру* түрінде ақпаратты сақтау мен ұсыну нысаны ретінде дұрыс мәліметтер (нысандар, процестер, құбылыстар және т.б. ғылыми, техникалық, экономикалық қасиеттер) көрініс табады. Құжаттық және фактографиялық жүйелер арасында айтарлық-

тай айырмашылық болмайды. Сондықтан аралас жүйелер ұғымын да атап кетуге болады. Мұндай жүйелерде құжаттық және фактографиялық ақпараттар бір ғана нысанаға қатысты болады.

Құжатты іздестіруде фактографиялық ақпарат қосымша мүмкіндік ретінде қолданылады. Фактографиялық элементтерге ие болатын құжаттық іздестірудің бір түрі ретінде *библиографиялық іздестіруді* атауға болады. Мұндай іздестіру тәсілі құжаттың тек библиографиялық элементтерінің баяндалуы бойынша ғана іске асады (шыққан жылы, басылым орны, басылым түрі, авторы және т.б.). Әдетте, құжаттық ақпарат-іздестіру жүйеде сақталатын мәліметтер рефератпен қоса есептелетін библиографиялық көріністе беріледі. Ондай жиымдарға жасалатын *сұрату бейнесі* көбінде фактографиялық сипатта болады.

Ақпарат іздестірудің негізінде құжаттағы ақпарат бейнесін сұратудағы ақпараттың бейнесімен теңестіру әдісі жатыр. Сондықтан олардың мазмұн көрінісін іздестіру жүйесіне «жіберу» мәселесін шешіп алу қажет болады.

Ол үшін құжат пен сұратудағы негізгі тақырыпты және олардың формалды сипаттамаларын баяндауға арналған мамандандырылған арнайы тіл — *ақпарат іздестіру тілі* құрастырылуы қажет. Қазіргі кездегі автоматтандырылған ақпарат іздестіру жүйелерінде қолданыс табатын барлық ақпарат іздестіру тілін мынадай үш типке бөліп қарастыруға болады:

1) ұғымдарды жүйелі түрде топтастыруға негізделген — ақпарат іздестіру тілінің топтастыру типі;

2) сөз немесе сөзтіркестер арқылы көрініс табатын лексикалық бірліктердің тізбесіне негізделген — ақпарат іздестіру тілінің дескрипторлық типі;

3) белгілі заттық аядағы нысандар «белгі-мағына» мазмұнындағы білім моделіне негізделген — ақпарат іздестіру тілінің нышанды-белгілік типі.

Құжат немесе сұрату мазмұндарына көрініс беру әрекетін **индекстеу** деп атайды. Ақпарат іздестіру тілі арқылы сипаттау (жазу) кезінде құжат мазмұны қорытынды түрінде және «қысылған» пішінде көрініс табады. Ақпарат іздестіру тілінде

жазылған құжаттың негізгі мазмұны құжаттың іздестіру бейнесі деп аталады. Ақпараттық мұқтаждықтың табиғи тілде жазылған пішіні — *ақпараттық сұрату* деп, ал оның ақпарат іздестіру тіліндегі көрінісі — *сұратуды іздестіру бейнесі* немесе *іздестіру ұйғарымы* деп аталып жүр.

Сонымен, ақпаратты іздестіру дегеніміз іздестіру жиымындағы іздестіру ұйғарымына сәйкес келетін құжаттың іздестіру бейнесін анықтауға әкеп соғады. Сәйкестіктің дәрежесін айқындау арнайы формалданған ережелер арқылы іске асады. Олар «шығарылу» (выдача) критерийі (өлшемі) немесе мағыналық сәйкестіктің критерийі (өлшемі) деп аталады. Бұл жерде тек формалды ережелер мен формалды релеванттық жайлы ғана сөз болып отырғандықтан, оны мағыналық релеванттылықпен ешқашан шатастыруға болмайды.

Шығарылу критерийінің ең қарапайым түрі ретінде құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің толық сәйкестігі немесе, дәлірек айтқанда, сұратудың іздестіру бейнесінің толық түрде құжаттың іздестіру бейнесіне қатысты болуы саналады. Бірақ, әдетте, іс жүзінде құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің жарым-жартылай сәйкестік критерийі (өлшемі) деп те қолданыс табуда.

Шығарылу критерийіне қойылатын талап, көбіне, оның тікелей іздестіру ұйғарымына қатысты болуы, сондықтан бір ғана ақпаратты іздестіру тілінің екі түрлі мүмкіндігі жайлы немесе бұл жерде құжаттардың және сұрыптаудың ақпарат іздестіру тілдері туралы айтуға болады. Бұл жағдайда сұрату тілі іздестіру мен шығарылу категориясына қатысты қосымша түрдегі арнайы мүмкіндіктерді өз құрамына қосып алады.

Ақпаратты іздестіру жүйесі үнемі адам мен машина аралығындағы жүйе немесе «адам-машина» жүйесі болып саналады. Сондықтан абстракты ақпарат іздестіру жүйесі (логика-семантикалық құрал) мен толық ақпарат-іздестіру жүйесінің (ақпараттық-іздестіру қызметі) түрлері деген атпен айырым табады. Осыған қарай, ақпараттық қызметтің: «сұрату — жауап» пен тандалып алынған ақпаратты тарату жолдары ажыратылады. Бұлардың біріншісі бойынша өткен кезді шолақамтитын жиым құжаттары негізінде бір реттік сұратулар алдын ала қарастырылады. Сондықтан мұндай іздестіру тәсілі, әдетте,

«ретро» (ретроспективный – *өткенді шолатын*) деп аталып жүр. Ал екіншісі бойынша – таңдалып алынған ақпаратты тарату жолында – жаңадан түскен құжаттар жиымы негізінде тұрақты (белгілі мезгіл ішінде ғана күші бар) сұратулар алдын ала қарастырылады. Орыс тілінде сұратудың мұндай тәсілі – «профиль» деп аталып жүр.

Адам мен ЭЕМ арасындағы «тілдесуге» қатысты «пакетті» және «диалогты» тәртіп жолдары өзара өзгешеленеді. Біріншісі, яғни «пакетті» тілдесу таңдалып алынған ақпаратты тарату тәсіліне тән болса, екіншісі – диалогтық тәртіпке – «ретро» тәсіліне тән. Пакеттік тәсілде іздестіру жолы адамның қатынасынсыз, әдетте, барлық сұратулар бойынша бір мезгілде орындалады.

Диалогтік тәсілде іздестіру кезіндегі аралық нәтижелер ескеріліп, қажетті мәселенің шешімі табылғанға дейін тұтынушы (немесе арнайы дайындалған маман) ЭЕМ-мен диалогке түседі, яғни сұрату мен іздестіру стратегиясын нақтылай түсетін шешім қабылдайды. Іздестірудің қай жолы болса да, тұтынушыға ұсынылған нәтиже – құжат, немесе екінші дәрежелі (библиографиялық баяндау, мүмкін, референтпен бірге), немесе бірінші дәрежелі қажетті құжат болуы мүмкін.

Ақпаратты іздестіру жүйелерін олардың функционалдық белгілеріне қарай, мынадай құрама бөлшектерге (ішкі жүйелерге) бөледі: құжатты енгізу, сұратуларды енгізу және өңдеу, іздестіру, сөздіктерді түзу, статистиканы анықтау, іздестіру нәтижелерін өңдеу және қажетті деген құжатты ұсыну.

Аталған қызметтерді іске асыратын құралдар мен әдістер *ішкі жүйелерді жабдықтау* деп аталып жүр. Әдетте, ішкі жүйелерді жабдықтаудың мынадай негізгі түрлерін атауға болады: *ақпараттық жабдықтау; лингвистикалық жабдықтау; техникалық жабдықтау; программалық жабдықтау; технологиялық жабдықтау.*

Ақпараттық жабдықтау – құжаттар мен сұратулардың ақпараттық жиымы және оларды сипаттаудың, құрастыру мен топтастырудың құралдары мен тәсілдері.

Лингвистикалық жабдықтау – ақпарат іздестіру тілінен, индекстеу ережелері мен ұсыну критерийінен тұратын логика-семантикалық аппарат.

Техникалық жабдықтау — ақпаратты сақтау мен іздестіруді қамтамасыз ететін техникалық құралдар.

Программалық жабдықтау — ЭЕМ арқылы ақпарат іздестіру жүйесінің барлық функцияларын іске асыратын программалық мүмкіндіктер.

Технологиялық жабдықтау — автоматтанған және автоматтанбаған процестер мен ақпарат-іздестіру жүйесінде ақпаратты өңдеу әрекеттерінің баяндалуы (описание).

Ақпарат іздестіру жүйені лингвистикалық жабдықтау дегеніміз ақпаратты іздестірудің негізгі міндетін орындауға арналған логикалық және лингвистикалық құралдар мен әдістердің жиынтығы. Лингвистикалық жабдықтаудың құрамы, ең алдымен, тілдік құралдардан тұрады — арнайы тілдер (оның ішінде шектеулі табиғи тілді қоса есептегенде), осы тілдерге аударудың ережелері және оларды қолдану әдістері. Ақпарат іздестіру тілін жүзеге асыру үшін әртүрлі сөздіктер, қалыптар (формулярлар) және лингвистикалық алгоритмдер құрастырылады. Лингвистикалық жабдықтаудың ұғымына ақпарат-іздестіру жүйесінде қолданылатын шығарылу критерийі, математикалық логика мен жиындар теориясының құралдары да бірге енеді.

Лингвистикалық жабдықтауға негіз болатын — *ақпарат іздестіру тілі*. Адамдардың қызметі жағдайында танымдық және коммуникативтік функцияларды атқаратын көптеген басқа да таңбалық жүйелерге қарағанда, ақпарат іздестіру тілі алғашынан-ақ формалданған ақпаратты іздестіруге арналған. Осы тұрғыдан ақпарат іздестіру тілінің айрықша ақпаратты-іздестіру қызметін сөз етуге болады.

Ең жақсы деген ақпарат іздестіру тіліне мынадай негізгі талаптар қойылады:

1) жеткілікті түрдегі «семантикалық күштілік», яғни құжаттар мен сұратулардағы негізгі тақырыпты (тақырыптарды) толық және дәл сипаттау мүмкіндігі;

2) бір мағыналық, яғни ақпарат іздестіру тіліндегі әрбір жазбаның (запись) бір ғана мағынаға және әрбір ұғымның ақпарат іздестіру тілі арқылы біркелкі түсінікке ие болуы;

3) ашық жүйе сипатында болуы, яғни тілдің лексикалық құралдарын үнемі толықтырып, өзгертіп, түзетіп тұру мүмкіндігі.

Кез келген жазба тіл, ең аз дегенде, үш түрлі компоненттен тұрады: әліпби, сөздік және грамматика. Бұл жағдай ақпарат іздестіру тіліне қатысты да орын алады. Әртүрлі ақпарат іздестіру тіліндегі лексикалық бірліктерді жазуға арналған графикалық белгілер жүйесі, яғни оның әліпбиі өз құрамында латын және орыс әліпбиін, араб цифрларын, тыныс белгілерін және арнайы символдарды сақтайды. Ақпарат іздестіру тілінің тілдік бірліктерінің жиынтығы *сөздік* деп аталады және тілдік бірлік ретінде ұғымдар мен ұғымдар топтары да бірге саналады.

Тілдік бірлік араларындағы қатынастарды анықтайтын жиынды *ақпарат іздестіру тілінің грамматикасы* деп атайды. Бұл қатынастар екі түрлі болып кездеседі: синтагматикалық (мәтіндік) және парадигматикалық (базалық). *Синтагматикалық қатынастар* дегеніміз нақты мәнмәтінде (яғни осы құжатта немесе сұратуда) белгілі бір жағдаятта пайда болатын тілдік бірліктердің араларындағы қатынастар.

Парадигматикалық қатынастардың туындауы үшін ақиқат әлемінің нысандары мен құбылыстары бейнелендіретін элементтердің орнына жүретін ұғымдар арасындағы логикалық байланыстың шарттылығы алынады. Табиғи тілдегі шағын ұғымдық түсінікте, грамматика дегеніміз қатынастардың бірінші типіне ғана қатысы бар мәселелер. Синтагматикалық қатынастарды белгілей алатын құралдары бар ақпарат іздестіру тілін «грамматикалы ақпарат іздестіру тілі», ал мұндай құралдары жоқ ақпарат іздестіру тілін «грамматикасыз ақпарат іздестіру тілі» деп атайды. Парадигматикалық қатынастар тілдің сөздік құрамына қатысты болады. Табиғи тілдердегі сөздіктерге қарағанда, ақпарат іздестіру тіліндегі сөздікте парадигматикалық қатынастар нақты түрде көрініс табады.

Координация ұғымына негізделген ақпарат іздестіру тілінің жалпы классификациясын сөз ететін болсақ, жоғарыда айтылғандай, ақпарат іздестіру тілінде тілдік бірлік ұғымдары, дәлірек айтсақ, ұғымдардың кластары орын алады. Ақпарат іздестіру тілі теориясында қарапайым ұғымдар кластарының қиылысулары арқылы құрастырылған күрделі ұғымдар кластарын

координация деп атап жүр. Аталған теорияда ақпарат іздестіру тілінің құрылымы мен индекстеу ережелеріне қарай, алдын ала координацияланған (предкоординированный) және соңынан координацияланған (посткоординированный) деген ұғымдар бар. Оның біріншісінде, құжаттарды индекстеу, оларға алдын ала құрылған күрделі кластарды тіркеп жазуға негізделеді де, ал ол кластардың аттары іздестіру бейнесіне айналады. Егер іздестіру бейнесі қарапайым кластардан тұрып, ал координация амалы (күрделі кластарды қарапайым кластардың қиылысуы арқылы құру) тек нақты іздестіру кезіне ғана сай келсе, ондай ақпарат іздестіру тілін *соңынан координацияланған* деп атайды.

Аталған жағдайларды ескере отырып, ақпарат іздестіру тілінің жалпы классификациясын мынадай түрде беруге болады:

1. Алдын ала координацияланған ақпарат іздестіру тілі.

1.1. Иерархиялық классификациялар.

1.2. Әліпбилі-пәндік классификациялар.

1.3. Фасетті (біржақты) классификациялар.

2. Соңынан координацияланған ақпарат іздестіру тілі.

2.1. Дескрипторлық тілдер.

2.2. Семантикалық тілдер.

2.3. Синтагматикалық тілдер.

Алдын ала координацияланған ақпарат іздестіру тілінде барлық кластардың аттары да алдын ала беріледі. Әрбір атауға қатысты табиғи тілдегі оған сәйкес келетін ұғымның атауы беріледі.

Ең көп таралған иерархиялық классификация. Мысал ретінде әмбебап ондық классификацияны, ГАСНТИ-дың рубрикаторын, ойлап табудың халықаралық классификациясын және т.б. атауға болады. Бұлайша классификация жасауға ұғымдарды бөлшектеу тәсілі негіз болып отыр. Классификацияда әрбір класс арнайы код арқылы белгіленіп, олар тұрақты және нақты орынға ие болады. Ақпаратты іздестіру тілі тұрғысынан қарастырғанда иерархиялық классификацияның негізгі кемшіліктері мыналар:

1) көпаспектілі индекстеуді қамтамасыз ете алмауы;

2) бөліктеу тереңдігінің жеткіліксіздігі;

3) ғылым мен техниканың дамуынан әрдайым артта қалып отыруы.

Иерархиялық классификациялау негізінен кітапханалар мен кітап қорларындағы алуан түрлі әдебиеттерді жүйеге келтіру жұмыстары үшін қолданылады.

Әліпбилі-пәндік және фасетті классификациялар нақты пәндік салада кездесетін негізгі категориялары мен пәндік кластардың алдын ала құрастырылған тізіміне негізделеді.

Әліпбилі-пәндік классификацияларда ұғымдар атаулары табиғи тілдегі сөздермен аталып, әліпбилік тәртіппен орналасады. Мұндай классификациялар шағын құжаттарды іздестіру жұмыстарына арналады. Бұлардағы негізгі тақырып сәйкес пәндік атпен белгіленіп, негізінен жүйелі каталогтерге нұсқағыш (көрсеткіш) құрастыру жұмысында қолданылады.

Фасетті классификация негізінде бір ғана жиындағы заттарды әртүрлі негізде топтау принципі жатыр. Дәлірек айтсақ, бір ғана жиын нысандары үшін бірнеше иерархиялық таралымдар құрылады. Фасетті классификациялау көпаспектілі индексация жүргізуге мүмкіндік тудырады. Негізгі категориялар — *классификациялау негіздемесі* — *фасет* деп аталады (ағылш. *facet* — *аспект*). Айталық, әртүрлі материалды белгілейтін терминдер жиыны «материалдар» фасетін құрайды, процестерді белгілейтін терминдер жиыны — «процестер» фасетін және т.б. құрайды. Фасеттің әрбір терминін *фокус* деп атайды және қолайлылық үшін оны қайсыбір код арқылы (цифрмен) белгілейді. Фасеттердің жиынын *классификация сызбасы* деп те атайды.

Мысалы, кинофильмдерді классификациялау үшін арнайы төрт фасетті бөліп алуға болады. Олар: жанр, метраж, түс және формат. Әрбір фасетте фильмді нақты аспект көзқарасында ғана сипаттайтын терминдер аталады. Сонда, кез келген фильмнің іздестіру бейнесі төрт фасеттен таңдалып алынған төрт термин-белгіден тұратын болады. Классификация сызбасында және оған сәйкес құжатты іздестіру бейнесінде фасеттердің орналасуы тұрақты белгіленеді. Мұндай орналасу жағы реттеліп, белгіленген тізбені фасеттік формула деп атайды. Сонымен бірге фасеттік классификациялау жұмысының үлкен қиындық туғызатындығынан, олар тек шағын пәндік сала үшін ғана жүргізіледі. Фасеттік классификацияны қолдану арқылы жүргізілетін индексациялау әдістемесі де аса күрделі. Сондықтан

автоматтандырылған ақпарат-ізвестіру жүйесінде аталғандай классификациялау іс жүзінде кең қолданыс таба алмауы мүмкін.

Сол сияқты семантикалық және синтагматикалық тілдердің де практикалық қолданысы жоқтың қасы. Бұлардың ішінен кең танымал болған Перри – Кенттің семантикалық коды, RX – кодының тілі, СИНТОЛ – тілі. Аталған тілдерге тән ерекшелік – ұғымдар мен айтылымдарды нақты сипаттайтын құралдың болуы. Ондай құралдардың орыс тіліндегі атаулары: «семантические множители, термы, реляторы, предикаты, сущности» және т.б.

Дескрипторлық тілдер. Автоматтандырылған ақпарат ізвестіру жүйелердің даму сипаты, негізінен, *дескрипторлық типті ақпарат ізвестіру тіліне* қатысты болады.

Дескрипторлық тілдерді құрастыруға координаттық индекстеу принципі негіз етіп алынады. Бұл ұстаным бойынша құжаттың негізгі тақырыбы – дескрипторлардың тізбесі, яғни қарапайым ұғымдардың аттарын білдіретін сөздер арқылы сипатталады. Мұндай сөздер қайсыбір ойша құрылған «n»-өлшемді пәндік-тақырыптық кеңістіктегі координаттардың орнына жүреді. Координаттық индекстеу принципі 50-жылдары механикаландырылған ақпарат ізвестіру жүйесін құрастыру жағдайында пайда болған. Мұндай жүйелердің ең алғашқыларының бірі деп АҚШ-та М.Таубе құрастырған УНИТЕРМ жүйесін айтуға болады. Бұл жүйеде құжат пен сұрату мазмұндарын баяндау индекстері ретінде мәтін ішінен тандалып алынған тірек сөздер (ключевые слова) қолданылады. Мұндай сөздердің (УНИТЕРМДЕР) барлығы да бірдей иерархиялық дәрежеге ие. УНИТЕРМДІК тілдердің сөздік құрамы алдын ала емес, тек индекстеу процесі кезінде тікелей құрастырылады. Шағын ұғымды анықтайтын координация немесе ұғымның логикалық көбейткіші екі реттен орын алатын тәрізді: құжатты индекстеуде – потенциалды (ойша) түрде және ізвестіру процесінде – ақиқаттық түрде. Ізвестіру процесі құжаттың ізвестіру бейнесін ізвестірудің алдын ала жазылып қойылған терминдерімен сәйкестендіру арқылы іске асады.

Мысалы, «кітапхана» және «автоматтандыру» ұғымдарының қиылысуы негізінде тар шеңбердегі жаңа ұғым – «кітапханаларды автоматтандыру» ұғымы туындайды.

Құжаттар мен сұратуларды тірек сөздер, яғни индекстеуге қажетті мәтіннен алынған терминдер арқылы индекстеу кезінде ақпарат іздестіру тілін лексикалық бақылау қажеттігі туындайды. Лексикалық бақылау – координаттық іздестіру кезіндегі сөздердің көпмағыналылығы мен біркелкі қолданбауының теріс әсерін барынша азайту мақсатында қолданылатын шаралар жиынтығы. Лексикалық бақылаудың негізгі мақсаты – мәтіндегі берілу тәсіліне қарамай, құжат пен сұрыптаудың іздестіру бейнелері ұғымдарының бірмағыналы түсінігін қамтамасыз ету. Бұл жердегі лексикалық бақылау термині *сөздіктік бақылау* сөз тіркесімен синонимдес. Сондықтан ақпарат іздестіру тілінде лингвистикалық бірліктерді нормаға келтіру үшін барлық маңызды тірек сөздер мен дескрипторлардың тізімі, ашық түрде берілген *индекстеу сөздігі* немесе *дескрипторлық сөздік* қолданылады. Парадигматикалық қатынастары қамтылған дескрипторлық сөздік – *ақпарат іздестіру тезаурусы* (грек thesauros – қор) деп аталады.

Тезаурус арқылы іске асатын лексикалық бақылаудың мәні негізінен мына төмендегіге саяды:

- 1) синонимдік сипатты бақылау;
- 2) омонимдік сипаттың басын ашу;
- 3) бір сөзге ғана қатысты оның барлық түрлерін (варианттарын) нормалау, яғни олардың бәрінің басын біріктіру.

Лексикалық бақылауды құраушылардың әрбіреуіне қысқаша тоқталайық.

1. Синонимдерді бақылау әрекеті ақпаратты іздестіру жүйесіндегі лексикалық бақылаудың ең негізгі қызметі деуге болады. Мұндай әрекеттің барлық синонимдері тек бір ғана лексикалық вариантпен ауыстырылады. Бұл вариант *дескриптор* деп аталады және дескрипторлық тілдің лексикалық бірлігі орнына жүреді. Дескриптор жиынтығы және оған мағыналық жағынан жақын жатқан сөздер шартты эквиваленттіліктің тобын құрайды. Синонимдер қатарына мыналар қосылады:

- толық синонимдер (радар – радиолакатор, ЭВМ – компьютер);
- сөздердің фонетикалық (графикалық) варианттары;
- аббревиатуралар (ақпарат іздестіру жүйесі);

– мән-мағыналық (смысл) жағынан бір ғана мағынасымен сәйкес келетін сөздер (компьютер – машина, мақала – жұмыс).

2. Лексикалық бақылау омонимдерді (омографтарды) және көпмағыналы сөздерді ажыратуды да өз міндетіне алады. Мұндай мақсат оларға әртүрлі кодтарды (шартты белгілерді) сәйкестендіру арқылы іске асады. Ақпарат іздестіру тілінде лексикалық бірлік сандық (цифрлық) әліпби арқылы жазылады немесе таңбалар (белгілер) жүйесі арқылы беріледі, мысалы: конденсаторлар (бу үшін) – конденсаторлар (электрлік), литье (предмет) – литье (процесс), меркурий (металл) – Меркурий (планета). Кейбір зерттеулердің нәтижелері бойынша омонимдерді ажыратпау ақпараттық жағынан көп кедергі (информационный шум) келтірмейді.

3. Мәтінде кездесетін сөзтұлғалар тірек сөздер ролін ойнап, индекстеу процесінде осы тірек сөздің (дескриптордың) сөздік формадағы нормаланған түріне (мысалы, атау септігі, жекеше) ауыстырылады. Кейде шартты эквиваленттіліктің бір ғана класына бір лексеманың грамматикалық формасына сайып қоймай, сонымен бірге әртүрлі негізді, тіпті әртүрлі сөз таптарына қатысты лексемалар ортақ лексикалық мағына негізінде (гиперлексема) біріктіріледі. Мысалы, сверло (тескіш) – сверление (тесу) және т.б.

Тезаурустың лексикалық бақылаудан кейінгі екінші маңызды қызметі – ақпарат іздестіру жүйесінің нәтижелігін, яғни толықтығын және дәлдігін арттыру. Бұған тек парадигматикалық және синтагматикалық қатынастарды қолдану арқылы қол жеткізуге болады.

Мәтіндегі лексикалық бірліктер арасындағы байланысты бейнелендіретін синтагматикалық қатынастарды белгілеу үшін арнайы грамматикалық құралдар, яғни рөлдер мен байланыс көрсеткіштері енгізіледі. Рөлдер көрсеткіші – қайсыбір дискриптор мәтінде (және оған сәйкес құжатты іздестіру бейнесінде) қандай мағынаға ие екендігін анықтайды және грамматикалық категориялар ролін атқарады. Мұндай категориялардың мысалы ретінде «процесті», «қасиетті», «материяны», «ортаны» атауға болады. Көрсеткіштер құжатты іздестіру бейнесіндегі сәйкес дескрипторларға тіркеп жазылатын кодтар арқылы белгіленеді.

Құжаттың іздестіру бейнесіндегі дескрипторлардың қайсылары өзара байланыста (логикалық немесе синтаксистік) болатындығын байланыс көрсеткіштері анықтайды. Олар да құжаттың іздестіру бейнесіндегі дескрипторларға тіркеп жазылатын айрықша символдармен белгіленген. Рөл мен байланыс көрсеткіштерін қолдану дәлдік дәрежесін арттыруға мүмкіндік туғызады. Рөл мен байланыс көрсеткіштерін пайдалану арқылы құжаттарды индекстеу қосымша еңбек сіңіруді қажет етеді. Одан кейін де неше түрлі қателер кездесуі мүмкін. Сондықтан, әдетте, рөл мен байланыс көрсеткіштері іс жүзінде қолданылмайды, ал олардың орны басқа тәсілдермен толықтырылады.

Парадигматикалық (базистік) қатынастар. Мұндай қатынастар бойынша ақпарат іздестіру тіліндегі лексикалық бірліктер арасындағы мән-мағыналар (смысл) байланысы көрініс табады. Аталған байланыс мәнмәтінге қатыссыз, ақиқат әлеміндегі нысандар арасындағы қатынастарға тәуелді болады. Бұларға иерархиялық қатынастармен бірге көптеген ассоциативті қатынастарды (бөлік – бүтін, зат – қасиет, процесс – нәтиже және т.б.) жатқызуға болады. Мән-мағыналық (логикалық) байланыстар нақты түрде (эксплицитті) тезаурус-тарда белгіленеді. Ақпараттық іздестіру тезаурусы – белгілі білім саласының мазмұндық тұрғыдағы семантикалық моделі.

Ақпараттық іздестіру тезаурусын құрастыру аса көлемді және күрделі жұмыстың нәтижесі деуге болады және оның негізгі кезеңдеріне жататындар:

- ақпараттық іздестіру тезаурусының тақырыптық аумағын қамтуды айқындау;
- мәтіндік деректерді жинастыру;
- тірек лексикалық бірліктерді жинастыру (сөзтізбені қалыптастыру).
- лексикалық бірліктерді шартты эквиваленттік кластарға топтау;
- лексикалық бірліктің араларындағы парадигматикалық қатынастарды анықтау және олардың типтерін таңдау;
- сөздік мақаланы және көрсеткіштерді (нұсқағыштарды) құру;
- ақпараттық іздестіру тезаурусын қалыптастыру.

– экспертиза мен тіркеу жұмыстарын орындау.

Ақпараттық іздестіру тезаурусын құрастыру кезінде қосымша құрал ретінде, әдетте, ЭЕМ қолданылады. Оның құрастыру ережесі, құрамы мен тұрпаты мемлекеттік стандартпен анықталады. Тезаурусты қалыптастыруда тұтынушы сұранысын қанағаттандыруға аса көңіл бөлінуі қажет. Оның тұрпаттық сипаты мағыналық жағынан қажетті лингвистикалық бірлікті және барлық оған қатысты лексикалық бірліктерді тауып алуды қамтамасыз ету көзделеді. Әдетте ақпараттық іздестіру тезаурусы екі бөліктен тұрады:

1. Дескрипторлардың әліпбилік тізімі және оның сөздік мақаласымен (ұясымен) қоса есептелетін сөздік бөлігі, басқаша айтқанда, ол тезаурустың өзі деуге де болады. Сөздік мақала дескриптордан басқаша, әдетте, ол төрт ішкі жиыннан тұрады:

– тірек синоним-сөздер;

– «жоғарыда орналасқан» («вышестоящие») дескрипторлар (осы дескриптормен «тек-түр» қатынасында тұратын).

– «төменде орналасқан» (нижестоящее) дескрипторлар (осы дескриптормен «тек-түр» қатынасында тұратын).

– ассоциативтік (ассоцияланған) дескрипторлар.

2. Тезауруспен іздестіруді жеңілдететін әртүрлі көрсеткіштер (барлық дерек сөздердің әліпбилік көрсеткіші, сөз тіркестерінің элементтері үшін пермутациялық көрсеткіш, жиіліктік көрсеткіш және т.б.).

Тезаурусты құрастыру жұмысында қиыншылық тудыратын жеке бір мәселе, ол – ендірілетін сөз тіркестерін таңдау.

Индекстеу процесі, яғни қол күші әрекетімен орындалатын құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесіне енгізілетін дескрипторларды таңдау – *индекстеу әдістемесі* деп аталатын арнайы нұсқаулықпен анықталады. Солай бола тұра, индекстеу аса тапқырлықты талап етеді. Іздестірудің нәтижелігі көптеген жағдайда индекстеудің сапалығына тығыз байланысты. Ал оның өз тарапынан, ақпараттық іздестіру тезаурусының және индекстеудің әдістемесінің сапалы болуына тікелей қатысы бар. Индекстеудің сапасын бағалауда іздестіру әрекетіндегідей индекстеудің толықтық (тереңдік) және дәлдік (егжей-тегжейлік) көрсеткіштері қолданылады. Ең алдымен, индекстеудің толықтық дәрежесі құжаттың іздестіру бейнесіне енгізілген

дескрипторлар санымен, ал индекстеу дәлдігі – олардың «сапасымен», дәлірек айтқанда, таңдалып алынған дескрипторлардың негізгі құжаттың (сұратудың) мазмұнына жақындық дәрежесімен анықталады. Бұл сипаттамалар өзара кері тәуелділіктегі қатынаста тұрады. Индекстеудің толықтық дәрежесі мен іздестіру жұмысын қамтамасыз ету үшін молынан индекстеу (избыточное индексирование) тәсілі қолданылады. *Молынан индекстеу* ұғымын құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесін негізгілерімен мағыналық жағынан тәуелді болып келетін дескрипторлармен толықтыру деп түсіну қажет.

Ірі көлемді құжаттар базасын құрастыру, ақпаратпен алмасу қажеттігі және іздестіру нәтижелігін арттыру мәселелеріне байланысты индекстеу әрекетін автоматтандыру проблемасы айрықша актуалдылыққа ие болды. Индекстеу процесін автоматтандыру мәселесін зерттеу ең алғаш Г.Лун (H.Luhn) және П.Баксендейл (P. Baxendale) және ресейлік Д.Г.Лахути есімдерімен тығыз байланысты.

Қазіргі кезде іске қосылған индекстеуді автоматтандыру жүйелері сапалық жағынан осыған ұқсас келетін автоматсыз түрде жасалатын интеллектуалды индекстеу жүйелерінен ешбір кем түспейді.

Жалпылама жағдайда, индекстеуді автоматтандыру сатысында сыртқы құжаттардан лингвистикалық бірліктерді тандап алып, оны құжаттың іздестіру бейнесіне енгізу мақсатымен нормалау процесі жүргізіледі. Лексикалық бірлікті тандап алу «оң сипатты» (мағыналы сөздер сөзтізбесі) немесе «теріс сипатты» (мағынасыз сөздер тізімі, «стоп»-сөз) сөздіктері бойынша орындалады.

Индекстерді автоматтандыру жүйесіндегі лексикалық бірлікті нормалау әрекеті морфологиялық талдау әдістері негізінде тілдік бірліктерді канондық түрге келтіру арқылы іске асады. Сонымен бірге, сөзтұлғаларды нормалау әдісінің тікелей іздестіру әрекеті кезінде де қолданылатын түрі бар. Нормалаудың ең көп тараған түрі (механизмі) – қию әдісі (немесе маскілеу). Бұл әдіс сұрату кезінде іздестіру термині ретінде тілдік бірліктің тек мағыналы бөлігі (негіз немесе квазинегіз) көрсетілетін жағдайда ғана орын алады.

Сапалы түрде индекстеуге қатысты қиындықтар тікелей рефераттар мен библиографиялық баяндаулар (описание) мәтіндері бойынша қажетті деген ақпаратты тауып алуды қамтамасыз ететін ақпараттық іздестіру жүйелерінің өмірге келуін тездетті. Ақпаратты іздестірудегі байқалатын ұстаным, ол іздестіру жүйесінің шектеулі табиғи тілге негізделген түрінің кең таралуы. Оларды тезауруссыз типті жүйе деп те атайды. Тезауруссыз ақпарат іздестіру жүйесінің кластары басқа жүйелерге қарағанда мынадай екі түрлі ерекшелікпен айырым табады:

- лексикалық бақылаудан бас тарту;
- құжаттарды индекстеуді автоматтандыру.

Мұндай ақпарат іздестіру жүйесінде координаттық индекстеу үшін табиғи тілдің сөздері, сөз тіркестері қолданылады. Шындығында, бұл жерде біз парадигматикасы мен сөздігі тұрақты түрде белгіленбеген дескрипторлық ақпаратты іздестіру тілімен әрекет етеміз. Ғалым А.И.Черныйдың пайымдауы бойынша, бұл жағдайда ақпарат іздестіру тілінің екі варианты қолданылады: біріншісі – табиғи тілдің тірек сөздері мен сөз тіркестерінің нөлдік қалпы, ал екіншісі – жоғары дәрежелі логикалық өндеу (сұратуды молынан индекстеу үшін) (*Черный, 1975*).

Аталған жүйелер УНИТЕРМ жүйесінен бастама алады. Мұндай тезаурустың қызметі мен қолданудағы ұтымдылығы жайлы жоғарыдағы баяндауымызда қысқаша айтылған болатын. Бірақ лексиканы бақыламайтын жүйелердің де өзіне тән құндылығы бар. Ең алдымен, бұған олардың аса дәлдік сипатын жатқызуға болады. Ешбір тезаурус табиғи тілдің лексикалық әралуандығын толығымен қамти алмайтындығы ақиқат шындық екені белгілі. Сондықтан индекстеу процесі құжат мазмұнын толық ашпайды, ал, тіпті, кейбір жағдайларда мазмұн өрескел бұрмаланады. Әдетте, көптеген жүйелерде тар сипаттағы көпаспектілі сұратулар басым болып келеді. Сол себепті табиғи тілге бейімделген жүйелердің үлкен артықшылығы бар, себебі, олар толық түрдегі ерекшелік сипатқа ие. Ал тезаурус ешқашанда ондай ерекшелікке ие бола алмайды. Сонымен бірге, әдетте, әртүрлі ақпараттық-іздестіру тезаурустары да бір-бірімен сәйкес келмейді. Кезінде ондай тезаурустардың өзара

сәйкес келу проблемасы ақпарат іздестіру жүйесі үшін негізгі мәселе болғанын да айта кетуге болады. Бұл мәселенің шешімін табу үшін әртүрлі тәсілдер қолданыс тапты, солардың ішінен аралық тілді құрастыру тәсілі де орын алған болатын. Бірақ бұл мәселенің шешімі табылмай, бұл проблеманы табиғи тілге бейімделген жүйелер толығымен күн тәртібінен алып тасталды.

Ақпарат іздестіру жүйелеріне баға беруге байланысты жүргізілген тәжірибелердің нәтижелеріне қарағанда, мәтін арқылы іздестіру жүйелері өзінің көптеген көрсеткіштерімен тезаурус типті жүйелерден қалыспайтындығын байқатты. Негізінен, ондай көрсеткіштерге дәлдік дәрежесінің ұлғаюы мен толықтық сипатының төмендеуі тән деуге болады. Бұл заңды да. Табиғи тіл арқылы іздестірудің ең жақсы нәтижелеріне жету үшін табиғи тілге тән «кемшілік» жақтарының орнын толтыру керек болды. Ол үшін ақпарат іздестіру жүйені екі түрлі мүмкіндіктердің (механизмдердің) жиынтығы деп қарастыруға болады: мағына ажырату және мағынаны сәйкестендіру. Бұл механизмдердің «күштілік» деңгейі бір-біріне сәйкес келулері керек. Айталық, егер біз ақпарат іздестіру жүйесінің мағына ажырату күштілігін арттыратын болсақ, міндетті түрде мағынаның сәйкестік механизмінің күшін де ұлғайта түсуіміз керек болады. Лексиканы нормалау мен лексикалық бақылау әрекеттері мағынаны сәйкестендіру қызметін жеңілдетеді.

Ақпарат іздестіру жүйесіне құжаттардың (рефераттардың) толық мәтінін енгізу құжаттың барлық лексикалық бірліктері мен олардың арасындағы мәнмәтіндік қатынастарды толық сақтап қалуға мүмкіндік туғызады. Байланыста тұратын мәтіндер бойынша әрекет ететін іздестіру жүйелерінің дәстүрлі дескрипторлық іздестіруден айырмашылығы – әрбір сөзтұлғаның мәтіндегі дәл орнын көрсету мүмкіндігі. Яғни бұл жердегі «орын» деп отырғанымыз сол сөзтұлға кездескен абзацтың нөмірі, сөйлемнің нөмірі, сөйлем ішіндегі сөздің нөмірі және т.б. Сөзтұлғалардың мәтін ішіндегі дәл орнын («адресін») білу сұрату мен іздестіру алгоритмінің арнайы мәнмәтіндік операторлар тілін туындатты. Аталған операторлар тілі сөзтұлғалар арақашықтығын және басқа да позициялық сипаттамаларға қатысты деректерді («позициялық операторлар») пайдаланады. Ал бұл жағдай лингвистикалық бірліктердің

синтагматикалық қатынастарын барынша жеңіл және икемді тәсілмен қолдануға мүмкіндік туғызады. Мұндай ақпарат іздестіру жүйесі – сызықтық грамматикалық жүйе деп аталады. Ондай жүйелерге ДИАЛОГ, CDS/ISIS, СКОБКИ, РАСПРИ деп аталатын қолданбалы программалар пакеттерін мысал ретінде келтіруге болады. Позициялық операторларды диалогтық іздестіру тәртібімен бірге қолдану жолы қарапайым әдіс-тәсілдер арқылы терминдердің жалған координатымен шатастырып алмауға мүмкіндік туғызады. Ал мұндай жалған сөзнұсқағыштық жағдай үлкен көлемді мәтіндер бойынша іздестіру жұмыстарын жүргізу кезінде сөзсіз болады.

Тезауруссыз ақпарат іздестіру жүйелерінде іздестіру ұйғарымын құрастыру сатысында мағынаны сәйкестендіру (және толықтықты арттыру) проблемасы өз шешімін табады. Себебі, мұндай ақпарат іздестіру жүйелерінде сұратуларды молынан индекстеу жұмысы іске асады. Онда іздестіру ұйғарымындағы әрбір терминге синонимдік немесе мағына жағынан жақын жатқан терминдер тіркеліп жазылып, ақпараттық іздестіру тезаурусындағы сөздік мақаланың толық пішіні орын алады.

Ақпарат іздестіру жүйесін лингвистикалық жабдықтаудың маңызды құрамдық бөлігі ретінде әрқилы сөздіктерді атауға болады. Олар автоматты немесе автоматсыз индекстеу тұсында, құжаттар мен сұратуларды жүйеге енгізу кезінде орын алатын қателерді автоматты түрде тауып алуда және т.б. жағдайларда қолданыс табады.

Мәтіндердегі қателерді автоматты түрде тауып алу әдістемесін екі үлкен топқа бөліп қарастыруға болады. Олар: бақылаудың полиграмдық (көпграфемалық) әдістері және сөздік арқылы бақылау әдістері. Бірінші жағдайда, қатені тауып алу құралы ретінде қарастырып отырған тілде мүмкін болатын екіәріптік немесе үшәріптік тіркестер алынады. Екінші жағдайда мәтін бойындағы сөздерге жасалған морфологиялық талдаудың нәтижесі эталон ретінде алынған сөздікпен салыстырылады. Соңғы әдістің, яғни сөздік әдісі бойынша іздестіру мен қателерді өңдеу жұмыстарының тиімді екендігі тәжірибе жүзінде дәлелденіп отыр. Бірақ жұмыстың нәтижелілігі сөздіктің көлемі мен талданатын мәтіндерді қамту толықтығына

байланысты болып келеді. Қатені табуға арналған ішкі жүйелердегі негізгі орын сөзөзгерту мен сөзжасам құбылыстарын ескере алатын автоматтанған морфологиялық әрекеттерге беріледі.

Ақпарат іздестіру жүйесін лингвистикалық жабдықтау жекелік сипаттағы маңызды бөлігі – машиналық сөздіктерді құрастыру мен іске қосуға қажетті құралдар мен тәсілдер. Әдетте мұндай мәселелер біріктіріліп, автоматтанған сөздік қызметі ұғымына енеді.

Шығарылу критерийі. Құжат пен сұратуды бір-біріне сәйкестендіргенде құжаттың сұратуға қарағанда релевантты болатындығын анықтау қажет болады. Осындай әрекеттің нәтижесінде сұрату бойынша тауып алынған құжаттың *шығарылуына* немесе *шығарылмауына* қатысты шешім қабылдау керек. Мұндай шешім – шығарылу критерийі арқылы қабылданады. Шығарылу критерийі басқаша атпен *мағыналық сәйкестік критерийі* деп те аталады. Сонымен, шығарылу критерийі немесе мағыналық сәйкестік критерийі дегеніміз құжат пен сұратудың формалды релеванттылық дәрежесін (яғни құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесі сәйкестігін) анықтайтын формалды ережелер деуге болады.

Дескрипторлық ақпарат іздестіру жүйесінде қолданылатын барлық шығарылу критерийлерін екі үлкен топқа бөліп қарастыруға болады. Олар: *есептелінетін* (вычисляемые) және *логикалық*.

Есептелінетін шығарылу критерийі сандық параметрлер бойынша әрекет етеді. Бұған құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің көлемге қатысты сипаттамалары да қамтылады. Бірқатар критерийлер олардың терминдер жиындары қиылысуларының теоретикалық-жиындық амалдарына негізделеді. Мұндай қиылысу шамасы пайыздық мөлшерде анықталуы мүмкін. Ең қарапайым жағдай – құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің толық сәйкес келуі. Бірақ мұндай сәйкестік аса қатал шарт түріне қатысты. Әдетте, тәжірибеде қиылысу шамасының қайсыбір «шекаралық» (пороговое) шамасы (мысалы, 50%) беріледі. Ал бұл шамадан жоғары сатыда тек релеванттылықтың деңгейі ғана тұрады. Шығарылу критерийінің қарапайым түрі құжаттың іздестіру

бейнесі мен сұратудың іздестіру бейнесіне енетін терминдердің бірдей дәрежедегі маңыздылығына келіп саяды. Бұл жағдайда шығарылу критерийінің шамасы мынадай өрнекпен (формуламен) анықталады:

$$R = \frac{N_i}{N_z} \cdot 100\%,$$

мұндағы N_i — құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесіндегі терминдерді іздестіру кезіндегі сәйкестік саны; N_z — сұратудың іздестіру бейнесіндегі терминдердің жалпы саны.

Сонымен қатар, көпшілік жағдайда әртүрлі терминдер, әсіресе іздестіру ұйғарымында, бірімәнді бола бермейді. Осыған байланысты терминге қатысты «салмақтылық» немесе «салмақ» ұғымы туындайды. Терминдердің салмағын анықтау амалы терминдерді «өлшеу» деп аталып жүр. Өлшеу тәсілдері әртүрлі болуы мүмкін. Олардың ең қарапайым түрі келесі пайымдауға негізделеді. Іздестіру ұйғарымын құрастыруда тұтынушы әрбір іздестіру ұйғарымындағы дескрипторды белгілі санмен «бағалайды». Міне, осы сан — *салмақ коэффициенті* деп аталады. Ол осымен қатар, тұтынушы салмақтар қосындысының «табалдырықтық» мәнін алдын ала көрсете алады. Сонымен бірге терминдерді «өлшеудің» автоматтанған әдістері де қолданыс табады. Бұл әдіске құжаттарды автоматты түрде индекстеу әдісін де жатқызуға болады. Мысалы, қайсыбір сөздің қайталанып қолданылуының қатынастық жиілігі қарастырылуы мүмкін. Осы құжаттағы сөздердің қатынастық жиіліктері барлық құжаттағы осындай сөздердің қатынастық жиілігінен асып түсу мүмкіндігі құжаттың негізгі мазмұны тұрғысынан аса маңызды деп саналады. Ал келесі бір тәсіл бойынша, құжаттың іздестіру бейнесіне қосу мақсатымен құжаттың ең маңызды деген тұстарынан (тақырыбы, түйіні және т.б.) таңдалып алынған терминдерге жоғары салмақтар тіркеліп жазылады.

Құжаттың іздестіру бейнесі мен сұратудың іздестіру бейнесінің сәйкестік дәрежесін анықтайтын шығарылу критерийі әртүрлі тәсілдермен есептеліп шығарылуы мүмкін. Жалпы жағдайда, әрбір «құжат — сұрату» жұптары үшін

векторларды салыстыру әдісінің көмегімен ұқсастық коэффициентінің шамасы есептеліп шығарылады. Осының нәтижесінде шығарылу құжаттарын, олардың ұқсастық коэффициенттерінің кему тәртібімен реттеуге және шығарылатын құжаттардың санын айқындауға мүмкіндік жасалады. Құжаттың саны алдын ала берілген «табалдырықтық» мәнге байланысты болады.

Қазіргі кездегі көптеген тәжірибелердің нәтижелері негізінде *ізвестіру коэффициенті* қолданыс табатын барлық ақпарат ізвестіру жүйелерінің жоғары тиімділігі дәлелденіп отыр. Бірақ іс жүзінде логикалық коэффициентке негізделген ақпарат ізвестіру жүйелері ізвестіру коэффициентіне негізделген жүйелерді бірте-бірте ығыстыра бастады. Мұндай жүйелер сұратудың ізвестіру бейнесі логикалық – «И», «ИЛИ», «НЕ» операторлары арқылы құрылады (орыс тілінің мағынасындағы сөздер $\&$, $\vee$, $\neg$ таңбаларымен белгіленеді). Осының нәтижесінде бір ғана ақпарат ізвестіру жүйесінің шеңберінде екі түрлі ақпарат ізвестіру тілі жайлы сөз етуге мүмкіндік туады. Оның екіншісін *булелік алгебраға* (булева алгебра) негізделген сұратудың арнайы тілі деп ұғыну керек. Мұндай жағдайда, егер сөз құжаттың ізвестіру бейнесінде бар болса, тірек сөздер немесе ізвестіру ұйғарымының дескрипторлары «1» («шындық») мәнін, ал ол сөз құжаттың ізвестіру бейнесінде жоқ болса «0» (нөл – «жалған») мәнін қабылдайды. Егер сұратудың логикалық формуласы, толығымен алғанда, осы құжатқа қатысты «шындық» мәніне тең болса, онда табылған құжат сұратуға релевантты, яғни олардың мазмұндары бір-біріне сәйкес келеді деп есептеледі.

Көпшілік жағдайда ізвестіру ұйғарымы конъюнктивтік нормадағы формалық (дизъюнкциялардың конъюнкциясы) түрде жазылады. Әдетте, синонимиялық қатынаста тұратын терминдерді «ИЛИ» (немесе) (дизъюнкция) операторлары арқылы байланыстырады. Ал «И» (мен) (конъюнкция) операторы арқылы сұратудың әртүрлі аспектілерін анықтайтын терминдерді (немесе топталған терминдерді) өзара байланыста ұстайды. **Булелік операторларды** қолдану құжаттар мен сұратуларды салыстырудың тұтынушыға түсінікті пішіндегі логикалық қисындылығын қамтамасыз етеді.

Сонымен бірге, логикалық шығарылу критерийі ескерілетін ақпарат іздестіру жүйелерде тұтынушыға ұсынылатын құжаттарды маңыздылығына қарай реттеу мен олардың көлемін бақылау қиындық туғызуда. Сондықтан тұтынушының өз қалағанындай нәтижеге қол жеткізе алмауына байланысты ақпарат іздестіру жүйесіндегі шығарылудың логикалық критерийінің «жеңісін» тек уақытша деп санауға болады.

Ақпараттық жабдықтау. Ақпараттық іздестіру жүйелерінде бірінші және екінші құжаттық «нұсқалар» ажыратылады. Бірінші «нұсқа» ретінде ЭЕМ жадында сақталатын іздестіру жиымдары (құжаттың іздестіру бейнесі, сұратудың іздестіру бейнесі) түсінік табады. Екінші «нұсқа» — әдетте, ЭЕМ-нан тыс (қағазда немесе микротаспаларда) сақталатын алғашқы құжаттардың жиымдары. Бірақ қазіргі кездегі ЭЕМ-дардың жадының ұлғаю сипаты толықмәтіндік деректер базалары санының көбеюіне әкеліп соқты да, соның нәтижесінде бірінші және екінші нұсқалық құжаттар бірігіп, біртұтастық көрініс тапты.

Қазіргі заманға сай ақпарат іздестіру жүйелері кең түрдегі ақпарат алмасу жиымдарымен (деректер базасымен) әрекет етуге арналатыны белгілі. Сондықтан олардың электрондық тұрпатының көрінісі арнайы халықаралық және мемлекеттік стандарттармен айқындалады, яғни басқаша айтқанда, олар тұрақты ережеге бағындырылады. Қазіргі кездегі ең кең тараған халықаралық стандарт — ISO 2709 (ГОСТ 7.14-84). Аталған стандартқа сәйкес жазылған барлық құжаттар мынадай бірыңғай құрылымға ие:

- жазу маркері — жазба бастамасында орналасқан және барлық жазбаға қатысты параметрлерді бойында сақтайтын, алдын ала белгіленген тұрақты ұзындықтағы өріс;

- анықтағыш — құжатта көрініс табатын деректердің әрбір библиографиялық бөліктеріне қатысты мазмұн элементтерінің жиынтығы;

- деректер элементтерін бойында ұстайтын деректер аясы (область данных).

Әрбір құжаттың толығы (деректердің библиографиялық бөліктерінің атаулары, олардың семантикасы, берілу тұрпаты) арнайы унификациялауды, яғни бірізге келтіруді қажет етеді.

Мұндай бірізділіктердің кең тараған түрі MARC ұясына жататындар (USMARC, UNIMARK және т.б.)

Техникалық және программалық жабдықтау. Техникалық және программалық жабдықтау – іздестіру қызметін (функциясын) іске асыру құралдары. Бұл жайтқа біз тек жалпылама түрде ғана тоқталамыз.

Ең алғашқы ақпарат іздестіру жүйелері адамның қатынасымен әрекет ететін және механикаландырылған жүйелер болатын. Қазіргі заманғы ақпарат іздестіру жүйелері ЭЕМ-ның негізінде іске асады. Жүйенің құрамындағы техникалық құралдарға қойылатын негізгі талап – олардың тез әрекет етуі мен жадының көлемділігі. Соңғы жылдары электрондық техниканың мүмкіндіктерінің барынша артуы көпшілікке мәлім, яғни қазіргі заманғы ЭЕМ-ның әрекет ету жылдамдығы (секундына миллиондаған амалдар атқарады) магнитті және оптикалық дискілер жадының көлемі *мега, гига және терабайт-тармен* өлшенеді. Практиалық тұрғыда қарастырғанда ЭЕМ-ның мұндай сипаттамалары кез келген іздестіру жүйесін құруға мүмкіндік туғызады. Ақпараттың шығарылуын қамтамасыз ететін неше түрлі «көру» мүмкіндігі бар терминалдық құралдар және әрқилы қағаз бетіне шығаратын құралдар (принтерлер) тез басу мүмкіндігімен бірге таңдау мүмкіндігіне де ие. Қазіргі ақпарат іздестіру жүйелері, әдетте, байланыс желілерінің және программалық-аппараттық комплекстердің негізінде ақпаратты қашықтықтан іздестіруді де орындай алады. Ақпараттың көлеміне және ақпарат іздестіру жүйенің мақсатына қарай іздестіру жүйелері ірі әмбебапты ЭЕМ-дар мен персоналды компьютерлер негізінде де құрастырылуда. Қажеттікке сай, персоналды ЭЕМ-дар (компьютерлер) үлестірілген базалар мен деректер банкілерін құруға мүмкіндік туындататын локалды желілерге біріктіріледі.

Ақпарат іздестіру жүйенің қызметін сүйемелдейтін программалық жабдықтауды шартты түрде екіге бөліп қарастыруға болады: жалпы жүйелік программалық жабдықтау және қолданбалы программалық жабдықтау. Қолданбалы программалық жабдықтау нақты ақпарат іздестіру жүйесінің барлық қызметін іске асырады: ақпаратты енгізу, бақылау, сақтау, индекстеу, іздестіру, шығару. Программалық құралдар-

дың құрамы және оларды іске асыратын әдістер көпшілік жағдайда лингвистикалық және ақпараттық жабдықтауға тәуелді болып келеді.

Ақпарат іздестіру жүйесінің программалық жабдықтау мәселесін зерттеуде деректер моделі мен ақпараттық базаны (іздестіру жиымын) құрастыру мәселелері маңызды орын алады. Деректерді ұйымдастыру тәсілін тандау әртүрлі факторларға тәуелді болып келеді. Мысалы, деректердің көлеміне, оларды қолданысқа енгізу жиілігіне, өзгерістердің сипаты мен динамикасына, техникалық құралдардың мүмкіндігіне, іздестіру режиміне (ИРИ, «ретро»), іздестіру логикасы мен шығарылу критерийіне. Деректерді ұйымдастыру тәсілдері ішінен деректерге еркін «кіру» әдістерінің бірізді және төтенше түрлері айырым табады. Бірізділікпен енудің (кірудің) негізінде, ол берілістерді біртіндеп, яғни бірінен кейін бірін алып өңдеу мүмкіндігі орындалады. Ал төтенше әдіс бойынша деректерге «қол жеткізу» үшін арнайы алгоритмдер құрастырылады және әдетте еркін тандалып алынған кез келген деректі тез табуға мүмкіндік жасайтын қосымша файлдар (деректер жиыны) құрастырылады.

Ақпарат іздестіру жүйесіне қатысты программалық жабдықтау мәселесін зерттеп, іске қосу үшін алгоритмдік тілдер мен операциялдық амалдар жүйелерінің құрамына енетін арнайы құралдар қолданылады. Іс жүзінде негізінен арнайы міндет жүктелген стандартты программалық комплекс қолданылады. Дәлірек айтқанда, құжатты ақпараттық іздестіруге арналған *қолданбалы программалар пакеттері* қолданылады. Мысал ретінде Ресей кеңістігінде көп тараған ІПП ДИАЛОГ және COS/ISIS қолданбалы программалар пакетін атауға болады. Бұл пакеттер, әдетте, іздестіру жүйесінің өзегін құруға мүмкіндік жасайды және мынадай құрамнан тұрады:

- деректер базасын құруға қажетті құралдар;
- деректер базасын толықтыру мен оған түзетулер енгізуге мүмкіндік жасайтын құралдар;
- сұрату мен іздестіруді құрастыруға қажетті құралдар;
- тұтынушы мен деректер базасы аралығындағы диалогтық қарым-қатынасты іске асыратын құралдар;

— ақпаратты шығару құралдары.

Аталған пакеттер тікелей «кіру» әдісі бойынша тез іздестіру үшін инверттелінген (дағдылы тәртібі өзгерген) файлдарды (индекстерді) құруға мүмкіндік береді. Инверттелінген файлдарды берілістің (деректердің) кез келген элементі бойынша құрастыруға болады. Ондай элементтерге мәтіндік өрістің сөзтұлғаларының да қатысы бар. Мәтіннің кез келген сөзі сол мәтіндегі «мекенжайымен» («адресі») бірге инверттелген файлда тіркеледі. Осының нәтижесінде, іздестіру кезінде сұратудағы терминдердің мәтін бөлігінің қажетті деген жеріне (мысалы, сөйлемге) дәлме-дәл түсуі үшін мәнмәтіндік (позициялық) операторларды қолдануға мүмкіндік туады. Бұл әсіресе толық мәтіндер үшін маңызды деп саналады. Ең бір дамыған толықмәтіндік ақпарат іздестіру жүйесі ретінде АРТЕФАКТ жүйесін атауға болады. Бұл жүйеде сөзтұлғаларды морфологиялық нормалау, деректерді «қысу» және сұратудың дамыған тілі сияқты әрекеттерді қолдану іске асады. Мәнмәтіндік операторлар іздестірудің сапалылығы мен икемділігін арттыруға маңызды сұратудағы сөз тіркестерін құруға мүмкіндік жасайды. Іздестіру мүмкіндігін арттыру үшін форматтық өріс бойынша іздестіру механизмін іске қосуға болады. Бұл тәсіл бойынша арифметикалық салыстыру операторлары (үлкен, кіші, тең және т.б.) арқылы қосымша шарттарды енгізуге жағдай жасалады.

Белгілі шектеуліктерді ескере отырып, форматтық өріс бойынша орындалатын іздестіру механизмі фактографиялық іздестіру кезінде де қолданылуы мүмкін. Солай бола тұра, фактографиялық ақпараттық іздестіру жүйесін, кейбір жағдайларда құжаттық іздестіру жүйесін де құру үшін *деректер базасын басқару жүйесі* қолданылады.

Автоматтанған фактографиялық ақпараттық іздестіру жүйесі — нақты деректердің (фактілердің), белгілі нысандардың қасиеттері және т.б. жайлы ақпараттарды сақтайтын ақпараттық жүйе. Автоматтанған фактографиялық ақпараттық іздестіру жүйені сипаттауда, әдетте, деректер базасы теориясының терминологиясы қолданылады. Себебі, ол жеткілікті түрде дамыған және автоматтанған фактографиялық ақпараттық іздестіру жүйені іске қосудың практикалық жағы қазіргі кезде

осы теорияның негізінде іске асуда. Деректер базасы ұғымын ЭЕМ жадында сақталатын деректер мен олардың арасындағы қатынастардың жиынтығы деп түсіну қажет. Белгілі бір мағынада деректер базасы «ақиқат болмыстың үзіндісінің» моделі немесе оның ақпараттық бейнесі деуге болады. Ал деректер базасында сақталатын құбылыстар, деректер (фактілер), заттар және т.б. жайлы ақпараттар нысан (объект) деп аталады. Басқа сөзбен айтқанда, нысан — біртұтас күйінде тұжырымдауға болатын деректер базасындағы ақиқаттықты бейнелейтін элемент. Мысалы, нысан ретінде ұшақтың рейсі (жолы), механизмнің бөлшегі, өндірістің қызметкері және т.б. алуға болады. Кез келген нысан әртүрлі қасиетке, сапаға ие болады. Міне, бұл аталғандарды нысанның *белгілері* немесе *атрибууттары* деп атайды. Атрибууттардың жиынтығы нысанды анықтайды да, ал ақпараттық көзқарас бойынша, нысан — осы атрибууттар жиыны деп саналады. Мәселен, *ұшақ рейсі* атты нысан мынадай атрибууттарды сипаттай алады: рейс нөмірі, ұшақтың типі, сапарға аттану пункті, сапардың аяқталу пункті, сапарға аттану күні, уақыты, сапардың аяқталу уақыты, ұшақтың әуеде болу уақыты, билет құны және т.б.

Егер программаларды жазу тілінің терминімен айтсақ, онда әрбір нысан бір «жазба» (запись) түрінде көрініс табады. Әрбір жазба атрибуутқа сәйкес келетін өрістерден тұрады.

Осындай терминологияға сүйенетін ақпараттық жүйелерді кейде *нысанды-белгілік ақпараттық жүйе* деп те атайды. Сонымен, деректер базасы — белгілі бір нысандардың жиыны (атрибууттар жиыны) және олардың аралық байланысы.

Деректер базасын құрылымдық тұрғыда қарастырғанда әдетте оны үш түрге бөледі: *иерархиялық*, *желілік* және *реляциондық*.

Деректік базаның құрылымы «шежірелік» (бұтақ таралымы) принципіне негізделсе, оны *иерархиялық құрылымдық деректер базасы* деп атайды. Егер деректік базаның құрылымы еркін желі құрылымындай болса — *желілік құрылымдық деректер базасы* болмақ. *Иерархиялық құрылымдағы деректер базасына* өңдеу жұмыстарын іске асыру үшін жан-жақты зерттелген математикалық құралдар кең түрде қолданылады. Ал желілік құрылымдағы деректер базалармен

әрекет ету соңғы кезде көптеген қиындықтарға соқтырып жүр. Сондықтан желілік деректер базасына қарағанда иерархиялық деректер базасы бойынша іздестіру жұмысын атқару аса нәтижелі болатындығы дәлелденді. Бірақ кез келген ақпаратты *иерархиялық құрылымда* бейнелеу ойдағыдай бола бермейді. Сол себепті егер ақпаратты баяндау тәсілі иерархиялық сипатта болса ғана, іздестіру, өңдеу жұмыстары иерархиялық деректер базасына қолданылуы керек. Соңғы кезде жиі қолданыс тауып жүрген реляциялық деректер базалар тікбұрыштық кестелердің жиыны түрінде көрініс табады. Ал кестелер арасындағы байланыс – айқыш сілтемелер арқылы іске асады.

Деректер базасы бойынша жұмыс істеу мүмкіндігіне ие болу үшін арнайы жазылған программалар қажет. Деректер базасындағы жұмысты қамтамасыз ететін программалар топтамасын *деректер базасын басқару жүйесі* деп атайды. Деректер базасын басқару жүйесінің құрамы әртүрлі деректер базалармен жұмыс істеуге мүмкіндік тудыратын барынша әмбебаптық сипаттағы программалардан тұруы қажет деп саналады.

Деректер базасын басқару жүйесі деректер базасының құрамдық жағын құрастыруды, деректер базасындағы ақпараттық өзгерісті (модификацияны), деректер базасындағы ақпаратты іздестіруді, тұтынушыға ақпараттың қолайлы түрде шығарылуы мен басқа да қосымша функцияларды қамтамасыз етуді көздейді. Осы аталғандарға толығырақ тоқталайық.

Деректер базасының құрамын құрастыру. Деректер базасын зерттеп, іске қосу барысында оның құрамдық жағына көп көңіл бөлінеді. Ақпаратты сақтау мен өңдеу нәтижелі болу үшін қандай ақпарат түрін сақтаудың қажеттігі, деректер базасында қандай нысандардың типтері орын алатындығы алдын ала айқындалуы қажет. Әрбір нысан типтері үшін оларды айқындайтын атрибуттар топтамасы анықталуы керек болады. Сонымен бірге әртүрлі типтегі нысандар арасында мүмкін болатын байланыс типтері де анықтауды қажет етеді. Барлық нысандардың типтері бейнеленген және әрбір нысан типінің атрибуттар топтамасы мен мүмкін болатын байланыс типтері көрініс тапқан сызбаны *жалпы концептуалдық сызба* немесе *деректер моделі* деп атайды.

Әдетте, деректер базасын зерттеу онда бейнеленетін ақпаратқа қатысты пәндік аяны зерттеуден және жалпы концептуалды сызбаны бейнелеуден басталады. Деректер базасын басқару жүйесінде жалпы концептуалдық сызбаның негізінде құрастырылатын деректер базасының құрылымы жайлы ақпаратты енгізу мүмкіндігі ойластырылуы керек. Осындай жағдайда әрбір атрибуттың деректер типін анықтау қажеттігі (программалау мағынасында) туындайды.

Әдетте деректер базасын құрастырғанда артықтық сипаттағы ақпараттың оның жадында сақталмағаны жөн деп саналады. Сондықтан, кейбір жағдайда, деректер базасын артық болмайтын (неизбыточная) деректер жиыны деп анықтайды. Бірақ іс жүзінде мұндай нәтижеге жету оңай емес. Көбінде ақпаратты іздестіру жұмысының ыңғайлы болуы үшін ақпараттың артықтығы қолайлы деп те саналады. Мысалы, *ұшақ рейсі* деп аталатын нысанның мынадай атрибуттары *ұшу уақыты*, *қону уақыты* және *әуеде болу уақыты* ішінен екеуі ғана белгілі болса, үшінші атрибутты әр уақытта есептеп шығаруға болады. Бірақ егер біз деректер базасынан ақпаратты іздестіру әрекетін *ұшақ рейсі* сипаттамаларының әрбіреуі бойынша жүргізу қажет деп санасақ, үш түрлі атрибуттың да орын алғаны жөн болады. Солай болғанның өзінде де, жалпылама жағдайда, ақпарат артықтығының аз мөлшерде болуына талпынуымыз қажет.

Деректер базасындағы ақпараттың артықтық сипатының қауіптілігі тек қосымша жад көлемін қажет етуде ғана емес, ең алдымен, деректер модификациясының өзгеріске ұшырау салдарынан қадағалауға қажетті қосымша бақылау әрекетінің қажеттігі болуында. Сондықтан зерттеушілердің тұжырымдауынша, деректер базасындағы артықтықтың тек бақылау жұмыстарына ғана қатысты болғаны жөн деп саналады.

Деректер базасындағы ақпаратқа өзгерістер (модификациялар) енгізу. Деректер базасын басқару жүйесінің деректер базасында орналасқан ақпаратқа тиісті өзгерістер енгізуге мүмкіндіктері болуы қажет. Ондай өзгерістердің типі мынадай болуы мүмкін: *жаңа жазбаны* (новая запись) немесе *жаңа нысанды қосу*, *бар жазбаны өшіру*, *бар жазбаның атрибут мәнін өзгерту*. Деректер базасының құрылымын

модификациялау ең күрделі өзгеріс түріне жатады. Іс жүзінде деректердің жалпы концептуалдық сызбасының идеалды түрін (өзгеріске жатпайтын) алдын ала болжамдау мүмкін бола бермейді. Сондықтан деректер базасын қолдану кезінде кейбір нысандар қажетсіз, ал қайсыбірінің жетіспеуі немесе артық болуы, тіпті кейбір тәуелділіктерді басқаша түрде ұйымдастыру қажет болады. Деректер базасын басқару жүйесінде деректер базасының құрылымын өзгерту тұсында ондағы қажетті деген ақпаратты сақтап қалу мүмкіндігі жасалуы қажет.

Ақпаратты іздестіру. Деректер базасын басқару жүйесі қызметінің ішінде ең бір маңызды деп саналатыны – тұтынушының сұранысы бойынша жүргізілген іздестіру әрекеттерінің нәтижелі болуын қамтамасыз ету. Іздестіру жұмысы дәл және тез уақыт аралығында орындалуы қажет деп ұйғарым жасалады. Ал ол үшін деректер жиынының белгілі пішінде ұйымдастырылуы қажет болады. Реттелген тізбе деректері бойынша іздестіру әрекетінің ойдағыдай нәтижелі болатындығы белгілі. Ал егер деректер тізбесі ретсіз түрде орналасса, онда қажетті деген нысанды тауып алу үшін барлық деректер жиынын толығымен, яғни басынан аяғына дейін қарастырып шығуға тура келеді. Мұндай әрекетке көп уақыт кетеді, ал уақытты үнемдеу үшін іздестіруді әртүрлі атрибуттардың мәні негізінде жүргізген жөн болады. Бұл мүмкіндік әдетте тек инверсияланған индекстік файлдарды құрастыру арқылы іске асады. Ондай файлдарда атрибуттардың мәні мен қажетті деген жазба түрлеріне сілтеме жасайтын арнайы түрдегі көрсеткіштер орын алады. Артықтық сипаттағы ақпаратты индекстік файлдар өз бойында сақтайды. Сондықтан мұндай ақпараттық артықтықты бақылау үшін деректер базасын басқару жүйесінде ақпаратты өзгертуге (модификациялауға) қажетті арнайы механизмнің орын алғаны жөн. Әрбір атрибутты олардың мәндері бойынша тез тауып алу үшін оның өзіне ғана сәйкес келетін, индекстік файлы болуы қажет. Яғни деректер базасын басқару жүйесінде ондай міндетті атқаратын арнайы программа (бағдарлама) құрастырылуы керек деген сөз. Мәні бойынша атрибутты тез іздестіру мүмкіндігін оның «кілтi» немесе *кілттік атрибуты* деп атайды. Жазбалар файлының реттелуіне қатысты «кілтті» *бірінші реттегі кілт*, ал

басқаларын *екінші реттегі кілттер* деп ажыратады. Әдетте *бірінші реттегі кілттің* өзіндік артықшылыққа ие болуына немесе жазбаның бірмағыналы теңестірулерге мүмкіндігі болуына талпыну керек болады. *Екінші реттегі кілттердің* мәндері әртүрлі жазбаларда қайталанып тұруы мүмкін. Сондықтан *бірінші реттегі кілт* бойынша іздестіруді ұйымдастырудың жылдамдығы аса жоғары, яғни ол максимум шамаға ие. Деректер базасын басқару жүйесі іздестіру жұмысын кілттік атрибуттардың мәні бойынша ғана емес, сонымен бірге кілтсіз атрибуттар негізінде де іске асыра алады. Бірақ соңғы жағдайда іздестіру жылдамдығы кемиді. Деректер базасында ақпаратты іздестіруге қатысты қарапайым екі түрлі типтегі сұратулар орын алады. Бірінші тип – айқын атрибуттың мәнін (шамасын) белгілі нысан бойынша (бірінші реттегі кілт) іздестіру. Сұратудың бірінші типіне мынадай мысалды келтіруге болар еді: «*Егер ұшақтың рейс нөмірі (бірінші реттегі кілт) белгілі болса, оның ұшу уақытын қалай табуға болады?*». Ал сұратудың екінші типі нысанды (немесе нысандарды) деп айқын атрибуттың белгілі мәні бойынша іздестіру әрекетін атайды. Мысалы: «*Егер ұшақтың қону пункті (жері) белгілі болса, оның рейс нөмірін (немесе рейс нөмірлерін) қалай табуға болады?*». Егер сұрату түрі күрделі болатын болса, ондайларды бірнеше қарапайым сұратуларға бөліп қарастыруға болады.

Деректер базасын басқару жүйесінің қосымша функциялары. Әдетте деректер базасын басқару жүйесі бірнеше қосымша қызметтерді қамтиды. Мысалы, ол көбінде іздестіру негізінде табылған ақпаратты тұтынушыға қолайлы түрде (мысалы, кесте түрінде) – шығарып беруді ұйымдастыра алады. Сонымен бірге деректер базасын басқару жүйесі көптеген есептеу жұмыстарын да орындай алады. Мәселен, мынадай сұратуды: «деректер базасында белгілі шарттарды алдын ала қанағаттандыратын қанша нысан бар?».

Деректер базасын басқару жүйесінде басқа да есептеу жүйелерінің атқаратын кейбір функцияларын өз міндетіне алатын мүмкіндігі болуы қажет. Мысалы, оған жататындар: ақпаратты жүйелік «жаңылысу» (сбой) кезінде пайда болатын қателерден, тұтынушының кездейсоқ қателерінен және

қажеттікке қарай кейбір жүйеге рұхсатсыз кіруден сақтандыру жұмыстары.

Деректер базасының теориясы автоматтанған фактографиялық ақпарат іздестіру жүйесін құрудың математикалық және программалық құралдары ретінде көрініс табатынын да айтуға болады.

Әдебиет

- Дейт К.* Введение в системы баз данных. М., 1980.
Информатика / Под ред. К.В.Тараканова. М., 1986.
Ланкастер Ф. Информационно-поисковые системы. М., 1972.
Мартин Дж. Организация баз данных в вычислительных системах. М., 1980.
Москович В.А. Информационные языки. М., 1971.
Панков И.П., Захаров В.П. «Информационно-поисковые системы» / В учебнике «Прикладное языкознание», отв. А.С.Герд, СПб., 1996.
Солтон Дж. Динамические библиотечно-информационные системы. М., 1979.
Сэлтон Г. Автоматическая обработка, хранение и поиск информации. М., 1973.
Тиари Т., Фрей Дж. Проектирование структур баз данных. М., 1980.
Черный А.И. Введение в теорию информационного поиска. М., 1975.

Автоматтанған оқу жүйесінің негізгі ұстанымдары

Оқытуды компьютерлендіру ісі қазіргі кездегі педагогиканың ең бір күшті бетбұрысы деуге болады. Сонымен бірге бұл білім беру жүйесінің алдында тұрған күрделі мәселелерінің біріне айналады. Тек оқытудың автоматтанған түрі ғана дәлелділік (уәжділік, мативация) сипатының жоғарғы деңгейінің негізінде оқып-үйрету әсерінің нәтижелігін түбегейлі өзгерте алады. Себебі білім беру барысының ескі (дәстүрлі) әдіс-тәсілдеріне жұмсалатын адам еңбегі (инемен құдық қазғандай) есептеу техникасына, яғни компьютерге жүктеледі. Сондықтан адам баласының бір рет қана атқаратын еңбегі оқып-үйренушінің қалауынша компьютер көмегімен сан рет қайталануына мүмкіндік туады. Әрине, оқытуды автоматтандырудан педагогтың еңбегін мүлде қажетсіз деген ұғым туындамауы керек. Бірақ автоматтандыру әдісі оқытушыны «қара жұмыстан» — шығармашылық әрекеттен босатып, оны қарапайым механикалық амалдарға ауыстырады. Ал мұндай жұмыстарды адам (мұғалім) әрекетімен салыстырғанда ЭЕМ (компьютер) «шаршамай-шалдықпай» зор нәтижелікпен және өте жылдам орындай алады.

Оқытуды автоматтандыру барысы, ең алдымен, автоматтанған оқу жүйесін жан-жақты зерттеп, құрастырып алып, тек содан кейін ғана оны қолдануды қажет етеді. Көбінде автоматтанған оқу жүйесі ұғымына оқытуға қажетті техникалық құралдар жиынтығы мен оқыту программаларының дестесін

(пакетін) жатқызады. Дестедегі әрбір бағдарлама тек белгілі бір сабақты жүргізуге, яғни оқытушы мен оқушы арасындағы әрекеттесудің бір ғана сеансын өткізуге икемделген. Мұндай тәсіл біржақты сипатқа ие, себебі «оқу жүйесі» ұғымы мынаған саяды: нақты автоматтанған оқыту (үйрету) программалары оқу жоспарына сәйкес келуімен бірге автоматтанған оқу жүйесінің ішкі құрылымдық көрсетімімен тұтастық кешенін құрайтын мағынада түсінік алуы керек. Мұндай тұтастықта қарастыру келесі екі негіздемеге сүйенеді: дидактикалық материалдың ортақтығы және оқытудың стратегиясы мен тактикасын автоматты түрде таңдау. Дидактикалық материал ретінде компьютерлік программаларды пайдаланатын белгілі пәндік саланың немесе кең мағынадағы білім базасының деректері қарастырылады. Екінші негіздік жайт — автоматтанған оқу жүйесі шеңберінде әрекет ету кезіндегі үйренушілердің алдыңғы қызметінде жүзеге асырылған мәліметтері, яғни оқу қызметі нәтижелерінің деректер базасы.

Дей тұра, автоматты түрдегі даралық оқыту программасының да белгілі дәрежедегі құндылық жағын теріске шығаруға болмайды. Бірақ автоматтанған оқу жүйесі құрамына енетін дара оқыту программасының нәтижелігі әр уақытта жоғары сатыда тұратынын ескеру қажет. Сондықтан ондай компьютерлік программалар даралық оқыту мен топтық оқыту процестерінің бір жүйе шеңберінде өзара әрекеттесу мүмкіндіктерін ескере құрастырылғаны жөн болады.

Жалпы жағдайда, автоматтанған оқу жүйесінің құрамы мынадай негізгі компоненттерден тұруы керек (*Жарков, 60*):

- 1) оқытудың техникалық құралдар кешені;
- 2) оқыту программаларының дестесі;
- 3) пәндік ая бойынша білім базасы;
- 4) оқыту әсерінің нәтижелерін тіркейтін деректер базасы.

Автоматтанған оқу жүйесі әртүрлі пәндерді оқыту барысында пайдаланылуы мүмкін. Оның қызметі сол жүйедегі программалардың оқушыға тигізетін әсері арқылы анықталады. Бұл программалардың негізгі қызметтеріне жататындар:

- оқытудың өзіне тікелей қатысты жағдайлар;
- жаттығу;
- бақылау;

— бағалау.

Автоматтанған оқу жүйесінің қызметін бұлайша бөліп құрастыру оқыту программаларын жобалау мен құрастыру кезінде аса маңызды: программа сценарийінің жалпы көрінісін, оқып-үйренушілердің диалог формасын (егер диалог алдын ала қарастырылса), сонымен бірге белгілі дәрежеде оқып-үйренушінің оқыту программамен өзара әрекеттесу сеансынан бұрынғы және кейінгі әрекеттерін анықтайды.

Автоматтанған оқыту жүйесінің негізгі ұстанымдарының бірі ретінде оның толық сипаты саналатын болғандықтан, яғни аяқталған оқыту үзіндісін жүзеге асыру мүмкіндіктерінде оқыту программаларының үш түрі де (ақпараттау, жаттықтыру, бақылау) толық сипатта орын алуы қажет. Тек осы аталған талап түгелімен орындалған жағдайда ғана автоматты түрде оқыту жүйесі толықтық сипатқа ие болады. Бірақ тек бір ғана оқыту программасы шеңберінде әртүрлі қызметтерді қамту алгоритмдерге жүктелетін мәселелердің негізгі айырмашылығына айтарлықтай күмән туғызады.

Жоғарыда аталған үш түрлі программалар үшін дидактикалық материалдар пайдалану тәсілімен айырым тапқанымен, бірдей дәрежеде қамтылуы қажет. Ал бұл жайт дидактикалық материалдардың (программада пайдаланылатын мәліметтердің құрылымы) өзара тәуелсіз болуының бірде-бір маңызды кепілі. Білім базасындағы немесе әмбебап тілде жазылған көрінісіндегі бірыңғай деректер өздерінің түрліше атқаратын қызметіне қарай, программалардың орындалу барысында әрқилы жүзеге асуы мүмкін. Мұндағы «әмбебап тілде жазылған көрініс» дегенді, деректер мен оларды өңдеу әдістерін формалды түрге келтіретін құралдардың жиынтығы деп түсіну қажет. Бірдей деректер қызметінің программадағы айырым табуы адам еңбегін үнемдеу ақиқаттығынан басқа, ол материалдарға (ұқсас және қайшылықсыз) өзгерістерді тек бір рет қана енгізуге мүмкіндік туғызады.

Ақпараттаушы (информирующие) программалар арқылы оқып-үйренушілер (обучаемым) дидактикалық материалдармен таныса алады, яғни ондай материалдардан тиісті ақпарат ала алады. Оқып-үйренушіге қажетті ақпарат беру тәсілі деректер жиынын айрықша безендіруді немесе оларды көрнекі түрде

экран бетінде орналастыруды, аса маңызды ақпаратты айрықша бөлектеп көрсетуді қажет етеді. Дұрысында, бір қалыптағы «жансыз» мәтіндік ақпаратты қажетті графиктер арқылы, компьютердің әуендік және басқа да құралдары арқылы оларға «жан бітіріп», оқып-үйренушінің жадында сақтау мүмкіндігін арттыру. Оқып-үйренушінің қалауынша бұрын танысқан ақпарат үзіндісіне қайта оралып, жаттығуына да жағдай жасалуы қажет.

Жаттығуға қатысты программалар өз құрылымына бөлек тәсілді қолдануды қажет етеді: оқып-үйренушіні білім базасына немесе оқу материалының ең қиын деген және біркелкі қате жіберілетін үзінділері бойынша түзілген жауаптар жүйесіне айрықша назар аударуға мүмкіндік туғызу. Сонымен бірге оқып-үйренуші мен программа әрекеті арасындағы (және керісінше) диалог жүргізілгені де жөн болады. Мұндай программаның орындалу барысында, оқып-үйренуші тарапынан кеткен қателерді атап көрсету, тіпті программаның күрделігіне қарай, ол қателердің не себепті орын алғанын да хабардар етіп отыру жөн саналады.

Бақылаушы программалардың жұмысы кезінде әрбір тапсырманың жауаптары деректер базасының оқыту барысы нәтижелерінде сақталып отыруы қажет. Сонымен бірге ақиқат және кездейсоқ қателерді аса ұқыптылықпен ажырата білу керек. Бұл жайт лингвистикалық оқып-үйрету программаларында аса маңызды деп саналады. Себебі жаңылыс басылған сөздердің салдары күрделі қателерге әкеп соғуы мүмкін.

Егер программада *бағалау* қызметі жүзеге асатын болса, қателерді әрі қарай да талдау қажеттігі туындайды, яғни кеткен қатенің «салмақ» деңгейін анықтау және орындалған тапсырма нәтижелерін статистикалық өңдеуден өткізу қажет болады.

Автоматтанған оқу жүйесінде бірнеше қосымша қызметтерді бөліп алып қарастыруға болады. Ондай қызметтерге төмендегілерді жатқызуға болады (*Жарков, 62*):

— оқу қызметі нәтижелерін тіркеу. Егер автоматтанған оқу жүйенің құрамында осындай деректердің базасы орын алса, онда мұндай қызметтің маңыздылығы арта түседі және оқып-үйренушілердің алған нәтижелерін тіркеу жұмысының әсерлігі де барынша өседі.

– оқып-үйренушіні ақпараттау қызметі. Егер білім базасының жүйесі арқылы оқып-үйренуші өз сұранысына қажетті деген ақпаратты алуына мүмкіндік туса, бұл қызмет толық көлемде жүзеге асады.

– диалогты ендіру қызметі. Диалогты пайдалану үшін жан-жақты зерттелген, қажетті деген айтылымдарды құрастыру мен тани білетін табиғи тілге барынша жақын келетін *жасанды тіл* құрастыру қажет.

– оқып-үйренуге қажетті стратегия мен тактиканы таңдау қызметі, яғни жүйені оқып-үйренуге икемдеу қызметі. Мұндай икемдеу қызметі үшін оқыту әрекеті бойынша құрылған деректер базасын пайдалану орынды деп саналады.

Аталған қосымша функцияларды оқыту программаларының түріне қарай үлестірілу сипаты да өзгеше болып келеді. Мәселен, нәтижелерді тіркеу қызметін жүзеге асыру тек бақылаушы программаларында ғана мүмкін, ал ақпарттау функциясының мұндай программаларға қатысы жоқ деуге болады.

Оқыту жүйесін құрастыруға байланысты бірнеше қолданбалы мәселелерді қарастырып, осындай қызметтерге қажет деп саналатын бірнеше ұстанымдар (принциптер) төмендегідей (*Жарков, 62-63*):

1) Көптеген автоматтанған оқу жүйесінің жоғары нәтижелігінің жеткіліксіздігі оқып-үйрену ісіне кешенді түрдегі қатынастың (подход) жоқтығы. Оқытудың кешенділік ұғымын ЭЕМ-ның (компьютердің) оқып-үйрену барысына толық араласу жағдайы деп түсіну қажет. Егер автоматты оқытудың компьютерлік программалары үзілді-кесілді сипатта болып, оқу курсының тек бірнеше бөлімдерін ғана қамтитын болса, ол нәтижелі бола бермейді.

2) Тілді оқытып-үйрету (обучение языку) көпаспектілі болуы қажет. Осыған байланысты оқыту программасын құрастыру кезінде оқып-үйрету мазмұнын құрайтын ақпаратты тілдің әр деңгейіне бірдей қатысты болуын мақсат етіп қойған жөн болады. Мұндай жағдайда оқыту барысының бірсарындылығы (монотондығы) азаяды. Осының нәтижесінде үйренушіде тек статикалық аспектідегі тіл жүйесі туралы түсінік қана қалыптасып қоймайды, сонымен бірге әртүрлі тілдік бірліктердің қызметі туралы да түсінік пайда болады. Мысалы, оқушы

жаңа атаулы (номинативті) бірліктерді меңгеру үстінде оларды бірден әртүрлі синтаксистік құрылымдарда да қолдануына мүмкіндігі бар. Осындай мәліметтермен бірге оқып-үйренуші атаулы бірліктердің валенттілік пен тіркесімдік қасиеттерін де сәйкес игере алады.

3) Оқыту барысының табиғилығы, біріншіден, үйретілетін тілдің ең ірі және жиі кездесетін құбылыстарын көрсете алатын оқу материалдарын мұқият түрде тандаудан тұрады; екіншіден, үйренуші жағынан басқаруды қажет ететін программаларда (мысалы, білім базасымен әрекет жасау) компьютермен жұмыс істеуде арнайы дайындықсыз-ақ диалог жүргізуді ұйымдастыру қажет болады; үшіншіден, компьютерлік программамен жүзеге асатын диалог тілі табиғи тілге барынша жақын дәрежеде жүргізілуі керек.

Үшінші жағдай қайсыбір нақтылауды қажет етеді. Көптеген диалог жүргізу мақсатындағы жүйелерде ағылшын тілі немесе соған ұқсас келетін тілдер қолданылады. Бірақ тілдерді оқыту программаларында, әсіресе, жаңадан үйренушілер үшін ана тілінің негіз болғаны жөн. Үйренуге (үйретуге) қажетті тіл ретінде, көбінде, ағылшын тілінің алынатынын ескерсек, оның интерфейсі ретінде, басқаша айтқанда, компьютер мен адам арасындағы диалогты жүзеге асыратын программалық құрал тілі ретінде қолдану үшін ағылшын тілін аздап болса да игеру қажет болады.

4) Табиғилық және вариативтілік принциптері бір-біріне іргелес жатады. Тіл жүйесі, сөйлеу тілінде әртүрлі тілдік құрылымдарды жүзеге асыру арқылы бір ғана немесе бір-біріне ұқсас келетін коммуникативтік міндеттерді көптеген тәсілдер арқылы шешуге мүмкіндік тудырады. Олардың қайсыбірін тандап алу, оқу барысында бейнеленетін экстралингвистикалық факторларға негізделеді.

5) Оқып-үйренуді автоматтандыру оқу барысында үйренушінің қажеттілік, қызығушылық деңгейін арттыруға мүмкіндік туындатады. Бірақ мұндай уәждік (мотивациялық) жоғары деңгейіне оқу программасының коммуникативтік негізде (яғни машина мен адам арасындағы диалогты жүзеге асыру арқылы) және оның жағдаяттық бағытталуы (бейімделгіштігі, яғни

тапсырманың жағдаяттығымен санасу) арқылы ғана жетуге болады.

Автоматтанған оқу жүйесін жобалау мен жете зерттеуде мынадай негізгі міндеттердің шешімін табу қажет болады:

- материалды тандау және оны топтау (әр сабаққа, тақырыптарға үлестіру);
- диалогтың сценарийін жете зерттеу;
- сыртқы пішін көрінісін безендіру немесе программаның дизайны;
- жүйедегі өзара әрекеттікті ұйымдастыру (оқып-үйретуге кешенді тәсілді жүзеге асыру);
- «деректер көрсетімі тілін» жете зерттеу;
- пәндік сала және онымен әрекет етуге қажетті интерфейс бойынша деректер базасын құрастыру.

Осындай әрбір ішкі міндеттердің әртүрлі оқу программаларының типтеріне қатынасы бірдей емес.

Автоматтанған оқу жүйесінің функционалды өзегін оқыту программалары құрайды. Себебі, үйренушінің оқыту программамен әрекеттесу барысы, шындығында да, автоматтанған оқу жүйесі арқылы жүзеге асады. Міне, осы әрекеттесу ғана оқыту әсерлігінің аса жоғары уәждігі (мотивация) мен нәтижелігін анықтайды.

Оқыту программалары әртүрлі негіздер жиыны бойынша топтастырылуы мүмкін. Олардың ең маңыздылары мыналар (Жарков, 64-66 бб.):

1) Оқып-үйренушінің ЭЕМ-мен (компьютермен) тікелей әрекеттесуіне оқыту программасының мүмкіндік тудыруына қарай. Кең мағынада алғанда, оқытуға қатысты программалар түр-түрге бөлінеді, бірақ олардың бәрі бірдей оқыту қызметін атқаруға бейімделмеген. Мәселен, ондай программалар қосымша сипатта мынадай қызметтерді атқара алады: оқу барысы жайлы ақпараттарды статистикалық немесе басқаша өңдеу қызметі, «әкімшіліктің» атқаратын көптеген қызметтерін есептеу құралдарына жүктеу (сабақтардың күнделік кестесін құру, оқып-үйренушілер мен оқытушылардың қызметтерінің нәтижелерін тіркеу және оларды техника жадында сақтау. Әсіресе, мұғалімдердің көп уақыты мен күшін қажет ететін оқу

бағдарламасын жасау әрекеті, есеп беру жұмысын атқару және т.б.

2) Оқыту процесінің аяқталған көрінісін жүзеге асыру үшін компьютерлік программа қандай қызметтерді атқаруына қарай: бақылау, қажетті әрекет түрін көрсету, оқып-үйренушінің әрекетін бағалау және т.б. Аталған міндеттердің әрбіреуін шешетін алгоритм тек оның өзіне ғана тән ерекшелікке ие, сондықтан олар бір ғана алгоритмдік тәртіп бойынша жүзеге аспайды; әртүрлі программалар үшін ортақтық сипат, ол оқыту барысына ешбір қатысы болмайтын кейбір таза формалды әрекеттер (процедуралар). Мысалы, терезелік интерфейс, жинақтауыштағы деректермен алмасу процедуралары, қыстырма (заставка) және басқа да программаның безендіру құралдары, оқып-үйренушімен диалог ұйымдастыру құралдары және т.б.;

3) Ұсынылатын оқу міндетінің типіне қарай: дағдылы (обычное), тапсырмаға ену (погружение); тапсырманың жағдаятына ену;

4) Оқу программасында диалогтың орындалуына және оның формасына қарай. Шындығында, автоматтанған оқу жүйесінің уәждік сипатын күшейтетін жеткілікті түрдегі маңызды құрал ретінде диалогты алуға болады; бірақ ол тағы да бір аса маңызды қызметке ие: бұл программа — оқып-үйренушінің «кері байланысын» (обратная связь) жүзеге асыратын бірден-бір құрал. Осының негізінде қандай да болсын формадағы күрделі оқу программалары диалогсыз жүзеге аспайды. Оқып-үйренуші жағынан да программаның орындалу барысын басқару және оның даралық ерекшелігін де есепке алу жұмыстары диалогсыз мүмкін емес. Диалог ескерілмеген оқыту программаларымен жұмыс жүргізу кезінде оқып-үйренуші әбден енжарлық күйде қалуға мәжбүр болады.

Диалог қайсыбір табиғи тілдің (ана тілі немесе ағылшын тілі) арнайы құрастырылған бөлігіне қатысты немесе жасанды бұйрықтар жүйесі негізінде құрылуы мүмкін. Анығы, оқыту программаларында табиғи тілді пайдалану қай жағынан алғанда да аса құнды, бірақ жасанды интеллект деңгейіне көтерілу қажет болатындықтан, ол көптеген қиындықтарға әкеп соғады; сонымен бірге оқып-үйренушілер бұйрықтардың арнаулы тілін алдын ала біліп алуы қажет болады және мұндай тіл тек бірен-

саран оқыту программаларында ғана қолданылып қоймай, оның бүкіл оқыту жүйесінде орын алғаны аса мәнді.

5) Оқыту программасының орындалу барысын оқып-үйренушінің өзі басқара алатын мүмкіндігіне қарай. Бұл жерде сызықтық (линейное) және тармақты алгоритмдер туралы сөз болып отыр. Бірінші жағдайда, яғни сызықтық алгоритмдерде программаның сценарийі бойынша оқып-үйренушілердің алдын ала берілген бақыланылатын және ұқсас әсерлер реттілігінен ауытқуға ешбір рұхсат жоқ. Екінші жағдайда, яғни тармақты алгоритмдер жағдайында аталған реттілік оқып-үйренушінің қалауынша құбылып тұруына мүмкіндігі бар. Бұл жағдай оқып-үйренушінің оқу барысына қатысының белсенділігін арттырады.

6) Оқып-үйренушінің даралық ерекшелігінің ескерілуіне қарай. Даралық ерекшелік дегенде оқушының өзіне ғана тән сапалық сипатынан гөрі, оқып-үйренушінің деңгейіне қатысты қайсыбір сандық сипаттама деп ұғынған жөн. Өзара айырым табатын деңгейлер саны барынша үлкен болуы да мүмкін (кейбір жүзеге асырылған автоматтанған оқу жүйесінде оның саны тоғызға дейін жетеді). Даралық ерекшелікті ескеру оқыту тәсілінің дербес болуын таңдаудан туындайды, яғни ол толығынан алғанда, оқу барысының нәтижелігін барынша ұлғайтуға мүмкіндік жасайды. Мұндай әр оқып-үйренушінің дара мүмкіндігін есепке алу, программаның іске қосылмай тұрған кезінде, яғни алдын ала және программаның орындалу барысында да ескеріледі. Соңғы жағдайда бір деңгейден екінші деңгейге өтуге мүмкіндік туады. Мұндай ауысу әрекеті аса қарапайым немесе аса күрделі тапсырманы таңдайтын оқып-үйренушінің қалауы бойынша да жүзеге асуы мүмкін және сонымен бірге программаның орындалу барысында алдыңғы әрекеттердің ойдағыдай жүруін автоматты талдау арқылы да жүзеге асыруға болады. Деңгейді бағалауда субъективтілікті (өзіндік сипатты) жою тұрғысынан автоматты талдау тәсілін қолдану ең ұтымды деп саналады;

7) Оқып-үйренушінің білім базасымен әрекет ету мүмкіндігіне қарай. Әрине, мұндай мүмкіндік білім базасымен әрекет жасауды жүзеге асыратын программаларды ғана сипаттай алады, яғни акпараттаушы, оқып-үйретуші және жаттық-

тырушы программалар бақылау жұмысына жарамсыз болып саналады.

8) Автоматтанған оқу жүйесінде программалардың орындалу нәтижелерінің тіркелу жағдайына қарай. Бұл мүмкіндік өзінің басқа да құндылықтарымен бірге нақты оқып-үйренуші мен автоматтанған оқу жүйесі арасында өтетін сеанстық жұмыстар әсерінің сабақтастығын қамтамасыз ете алады. Мысалы, әрбір оқып-үйренушіге қатысты қайсыбір программаның орындалу жағдаяты тіркеледі. Ал бұл оқыту барысында кездесіп тұратын кейбір тақырыптың сырт қалуынан және қайталануынан сақтандырады. Сонымен бірге нәтижелерді тіркеу ісі оқып-үйренушінің білім деңгейін оқу барысында ескеріп, реттеп отыруға мүмкіндік тудырады.

9) Дидактикалық материалдардың программа мәтінінде немесе оған тәуелсіз деректер берілісінде орын алуына қарай. Мұндай жағдайда әртүрлі деректер жиынтығымен бір программа арқылы әрекет етуге мүмкіндік туады. Осының нәтижесінде оқыту жүйесін программалауға жұмсалатын адам еңбегі мен оқу-үйрену материалдарына тиісті өңдеулер жүргізу уақыты үнемделеді.

Осы айтылғандарға байланысты автоматты оқу жүйесінде көптеген міндеттерді шеше алатын және икемділігі мол деректер көрсетімі тілін (язык представления данных) зерттеп құрастырудың маңыздылығы арта түсуде.

Ә д е б и е т

Автоматизированные обучающие системы на базе ЭВМ / Под. ред. А.Ф.Чернявского. Минск, 1980.

Жарков И.В. Автоматизированные обучающие системы // Прикладное языкознание: Учебник / отв. ред. А.С.Герд. СПб.: Изд-во С.-Петербург. ун-та, 1996. 59-67-бб.

Носенко Э.Л. ЭВМ в обучении иностранным языкам в вузе. М., 1988. 103 б.

Пиотровский Р.Г. Компьютеризация преподавания языка. Л., 1988.

Рыбаков Ф.И., Соколова И.Н., Лоскутникова Л.В. Программированное обучение с использованием ЭВМ. М., 1968.

Савельев А.Я., Новиков В.А., Лобанов Ю.И. Подготовка информации для автоматизированных обучающих систем. М., 1986. 174 б.

Фадеев С.В. ЭВМ в преподавании русского языка как иностранного. М., 1990. 81 б.

Автоматты (машиналық) аударма

Сөйлеу әрекетінің қайсыбір түрі қайта туындалатын автоматты жүйелер кибернетикалық модельдер тобына жатады.

Жалпылама айтқанда, адам баласының сөйлеу әрекетінің негізгі үш типін бөліп қарастыруға болады. Ал оларды түгінұсқаға барабар туындату үшін кибернетикалық модельдердің де ең кем дегенде үш типі құрастырылуы қажет:

1) коммуникация кезіндегі адамдардың тілдік әрекеті. Бұл жердегі коммуникация процесін кең түрде ұғынуға болады, яғни адамдардың бір-бірімен тікелей тілдесуін ғана емес, сонымен бірге белгілі ортада – баспа, радио, теледидар және т.б. арқылы тілдесуін де қамтиды. Осы типке қатысты тілдік әрекеттің негізгі формасы – әртүрлі эксперттік жүйелерде модельденетін диалог түрі;

2) қайсыбір білім саласын оқыту жағдайындағы, тілге үйретуді қоса есептегендегі тілдік әрекет. Мұндай әрекет типі оқыту жүйелері арқылы модельденеді;

3) шет тіліндегі мәтінді ана тіліне және керісінше аудару тұсындағы адамның тілдік әрекеті.

Машиналық аударма жүйелері кибернетикалық модель жасау аясына қатысты болуы себепті тілдік әрекеттің үшінші типі арқылы іске асады. Сондықтан аударма ісі электронды-есептеу машиналарының (ЭЕМ) көмегімен аудармашы әрекетін қайталауға бағытталады.

Аудармашының тілдік әрекетінің, яғни сөйлеу-ойлау қызметінің моделін машиналық аударма жүйесі арқылы жүзеге асыру екі түрлі таптастыру параметрлері бойынша анықталады:

– автоматтандыру дәрежесі, яғни компьютерге жүктелетін жұмыстың нақты көлемі. Мұнда мынадай жағдаяттарды айырып қарастыру керек (Кулагина, 1979, 3):

1) аударма жұмысы толығымен немесе толыққа жуық ЭЕМ арқылы орындалады;

2) аударма жұмысын аудармашы атқарады да, ал ЭЕМ жәрдемші рөлін атқарады (мысалы, қажетті сөздің аудармасын сөздіктен тауып алу жұмысында).

– аудармашы адамды қатыстыру стратегиясы, яғни аударма жұмысының негізгі бөлігі компьютер арқылы орындалады да, ал аудармашы аударылған мәтінді өңдеуге, яғни аудару кезінің әр тұстарында (басында, ортасында, аяғында) редакциялық жұмысқа қатынасады (*Machine Translation*, 1987, 12-13).

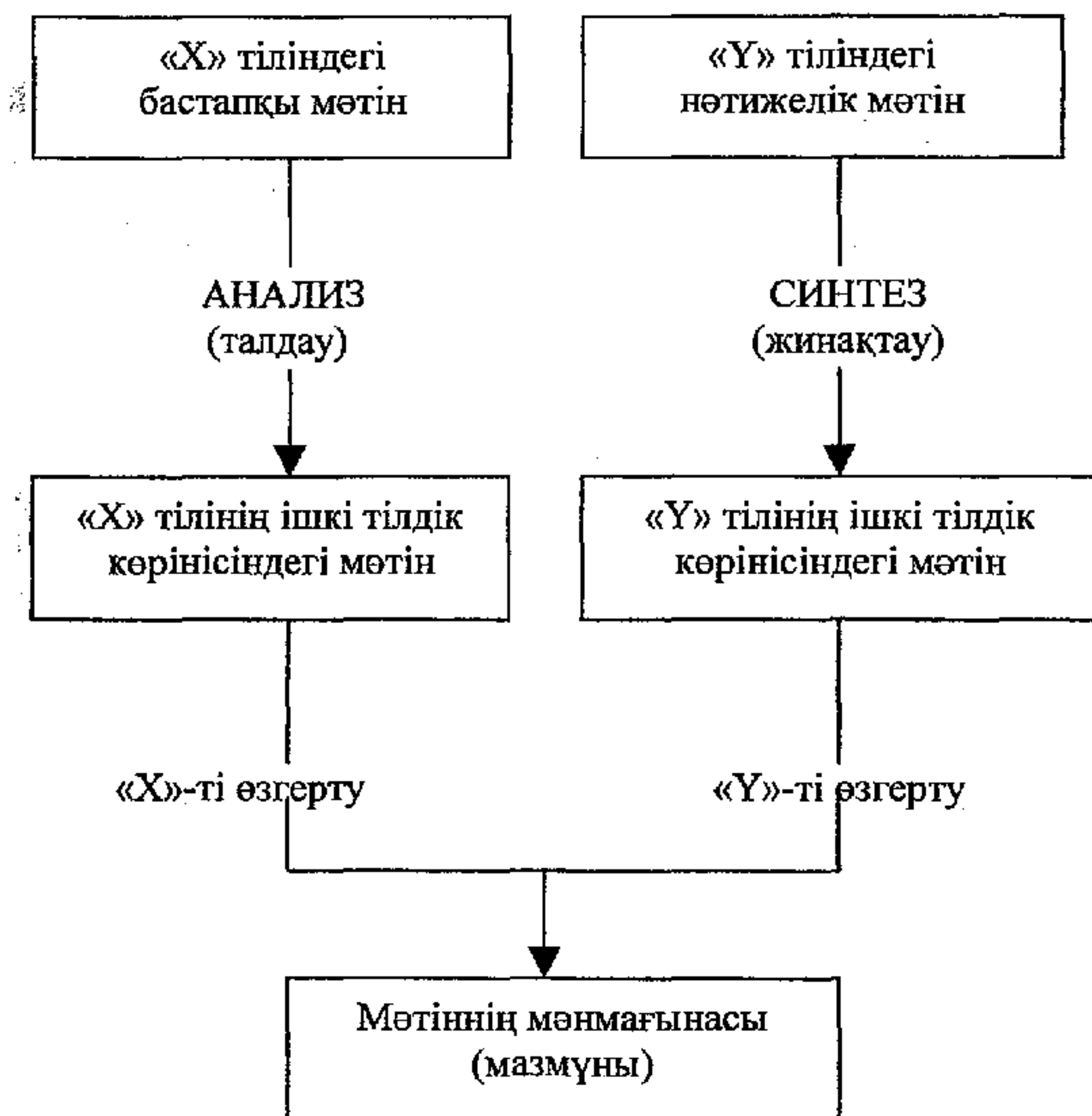
Аудармашының сөйлеу әрекетінің ерекшелігі мынаған саяды: бастапқы нысан ретінде бір тілдегі мәтін алынады да, ал тілдік қызметтің нәтижесі ретінде – мағыналық жағынан аударылуға тиісті мәтінге барабар, басқа тілдегі мәтін болып табылады. Басқаша айтқанда, аталған нәтиже міндетті түрде мәтін түрінде ұсынылуы қажет. Ал басқа бір тілдік әрекет типтерінде, мысалы, бұлшық ет күші арқылы немесе ойлау қызметі, білімді қорландыру және т.б. әрекеттерде нәтиженің мәтін түрінде болуы міндетті емес.

Аталған ерекшелік автоматты аудармаға белгілі принциптік құрылымды сақтауды міндеттейді. Төменде ондай құрылым мүмкіндігінің ең жақсы деген (идеалды) көрінісі берілді.

«Х» тіліне талдау жүргізу мен оны өзгертудің мақсаты – «Х» тілінің сөздігі, грамматикасы және қолдану ережелері негізінде «Х» тіліндегі мәтіннің мазмұнын ашу. Ал «Ү» тіліне жинақтау әдісін қолдануды – «Ү» тіліндегі шығару мәтінін құрастыру (немесе оны синтездеу) деп атайды. «Ү» тіліндігі мән-мағына сол тілдің сөздігі мен грамматикасы негізінде ашылады.

ЭЕМ-сыз дәстүрлі аудармашының мәтінді аудару кезіндегі талдау мен жинақтау компоненттерінің орын алуы айқын түрде

байқалмағанымен, **автоматты аударма жасау процесінде ондай тәсіл бөліктері міндетті түрде орын алуы қажет.**



*Сызба. Автоматты аударма процесінің
идеалдық құрылымы*

Ал сызбадағы «өзгерту» топтамасының қажеттілігі ең қарапайым жағдайда да айқын байқалады. Мысалы, аударуға қажетті «X» тілінің сөздік бірлігі мағыналық жағынан екінші тілдің («Y») сөздік бірлігімен және екі тілдің («X» пен «Y») синтаксистік құрылымдары да бір-бірімен сәйкес келіп тұрған жағдайда да (тіпті сөздердің орналасу тәртібіне дейін) тиісті өзгерістердің жасалатыны белгілі. Негізінен алғанда, сөздердің грамматикалық тұлғалары мен аударма баламасының сөйлем ішінде дұрыс орналасу тәртібі өзгертуді қажет етеді.

Бірақ жоғарыда қарастырылған жайт идеалды жағдаятына сай және тілдің шектеулі сөздік бөлігі мен синтаксисіне тән болып келеді. Сонымен бірге мұндай жағдаят тек әрбір нақты тілдер жұбы үшін ғана ұтымды деуге болады. Әзірше екі тілдің лексикалық бірліктері мен синтаксистік құрылымдары бойынша басым көпшілік тілдерге жоғарыда аталғандай тұспа-тұс сәйкестік жайлы айту қиын. Оған себеп – әртүрлі тілдердегі көп мағыналы сөздер көлемінің сәйкес келмеуі. Лексикалық сәйкестіктің ең соңғы бұзылу жағдайы – қайсыбір тілдерде баламасыз лексиканың болуы. Әдетте, олар тұрмыстық бұйымдарды, мәдениетті сипаттайтын сөздердің екінші (аударуға қажетті) тілдің лексикалық қорынан орын алмайды. Мұндай жағдай лексикалық элементті сипаттау немесе калькілеу үшін оның мәнін дұрыс түсінуді қажет етеді. Басқаша айтқанда, мағынаға сәйкес келтіру үшін қажетті деген өзгертулерді (өңдеулерді) жүргізу керек болады.

Егерде аударма мәселесін, *автоматты аударманы* қоса есептегенде, тіларалық сәйкестіктің бұзылуы тұрғысынан қарастырсақ, онда оның мақсатын екі тіл жұбы мәтіндерін мағыналық сәйкестікке келтіру деп ұғыну қажет. Яғни екі тіл арасындағы әртүрлі элементтердің (сөздік, құрылым, сөздер тәртібі және т.б.) мазмұнды ашу құралдарын сәйкестікке келтіру деп түсіну керек болады. Мұндай сәйкестікке келтіру мәселесін шешу жоғарыда көрсетілген сызбадағы топтамалар қызметінің барлық машиналық аударма жүйесінде орындалуын қажет етеді. Сәйкестікке келтіру мүмкіндігі және аударманың принциптік жақтарының барлық тілдерге ортақ болуы тілдің сызықтық және таңбалық сипаттарына негізделеді (Ф. де Соссюр). Ал таңбалық қасиеттің негізінде (мағыналық тұрғыда) барлық халықтардың ойлау қызметін таңбалау мен қоршаған ортаның физикалық ортақ қасиеттерінің жататыны белгілі.

Машиналық аударманың өмірге келуіне, біріншіден, XX ғасырдың екінші жартысынан бастап әр елдерде (континенттерде) бірнеше тілдегі ақпарат ағымының қарқындап өсуі, екіншіден, ғылыми-техникалық прогресс үшін оларды меңгеру қажеттігіне қатысты әлеуметтік себептер негіз болды. Осыған байланысты машиналық (автоматты) аударма жұмысының даму барысына қатысты кейбір жайттарға тоқталайық.

Аталған әлеуметтік себептер аударма мәселесінің шешімін табуды және оның ұтымды жолдарын іздестіруді қажет етті. Бұған ең алдымен аудармашының тілдік қызметінің ерекшеліктері ішінен шеттілдік мәтіннің құрылымын зерттеудегі таза формалды талдау жолын жатқызуға болады (*Панов, 1958*). Мысалы, аудармашыға мәтіннің грамматикасы және лексиканың көмекші сөздері мен есімдік сөздері және басқа да осы сияқты функционалды грамматиканың элементтері белгілі болғанымен, салалық мәтіннің арнаулы лексикасы таныс болмауы мүмкін.

Шеттілдік мәтінді аударудағы оның тұрпаты (формасы) бойынша мән-мағынаны ашу ерекшеліктерін модельдеу, ең алдымен, аудармашының іс-әрекетін зерттеу нысаны ретінде алуға бірден-бір себепші болды. Сонымен, мұндай модельдеу шеттілдік мәтінді аударуда тілдің ақиқат деректерін «шеттетіп», тілдің семантикасын арнайы талдамай-ақ, формалды тұрғыда аударма баламасын табуға мүмкіндік жасайды.

Автоматты аударма жасаудың бастамасы ретінде ресми түрде 1949 жыл саналады. Тілді формалды тұрғыда қарастыра отырып, электронды-есептеу техникасының көмегімен аударманы жүзеге асырудан бұрын оны механикалық машиналар жолымен іске асыру идеясы аталған мезгілден де көп бұрын бастама алғаны белгілі. Мысалы, А.Кахердің ойлап тапқан аударма-жазба машинасы жайлы мәлімет «Ұава таа» газетінде 1924 жылдың 24 ақпанында шағын хабарлама ретінде жарық көрген болатын (*Переводная..., 1962*). 1933 жылы Кеңес Одағында П.П.Смирнов-Троянскийге бір тілден екінші тілге аудару машинасын ойлап шығарғаны үшін арнайы патент тапсырылғаны да жұртшылыққа мәлім (*Переводная..., 1959*).

Аударма жасауда екі немесе одан да көп тілдердің бірліктері арасындағы мағына сәйкестігін анықтау үшін алғашында механикалық тәсілдер қолданылды. Сондықтан аударма жұмыстарының жетістіктері әлемдік ғылымға тек XX ғ. 50-жылдарында ғана белгілі бола бастады. Ең алғашқы машиналық аударма жүйелері АҚШ-та пайда болды. Тарихи тұрғыдан қарастырғанда машиналық аударма жүйесін құрастыруға жалпытілдік көзқарастағы криптография (шартты белгілер-

мен жазылған құпия жазу) саласының теориясы мен практикасы негіз болды.

Екінші дүниежүзілік соғыс кезінде әр елдерде, әсіресе Англия мен АҚШ-та құпия хаттардағы шифрлы жазуды үйреншікті жазуға ауыстыру (дешифровка) жұмысын машинасыз да және машинамен де әрекет ету арқылы көптеген нәтижелерге қол жеткені белгілі. Мысалы, гитлерлік басқару орындарының шифрланған бұйрықтары мен хабарларын жай жазуға айналдыратын «Ультра» атты дешифровалды машинаның құрастырылғанын айта кетуге болады.

Лингвистикалық тұрғыдан қарастырғанда, шифрланған (кодталған) хабарды дәстүрлі жазуға келтіру дегеніміз, ол лексикалық қоры белгісіз және грамматикалық ережелері әлі анықталмаған шет тілінде жазылған мәтінді аудару жағдаятымен барабар деуге болады. Мұндай аударма жұмысы бір кодтағы жүйеден екінші кодтағы жүйеге ауыстыру әрекетімен бірдей келеді.

Маман-дешифровщиктердің пайымдауынша, шифрланған (кодталған) хабарды жай жазуға келтіру әрекетіне қарағанда бір табиғи тілдегі мәтінді екінші табиғи тілге аудару әрекеті жеңіл деп бағаланады. Себебі, қазіргі табиғи тілдерді сипаттайтын көптеген сөздіктер мен грамматикалардың негізінде тілдің өзіне ғана тән лексикалық, морфологиялық және синтаксистік ерекшеліктері алдын ала белгілі деуге болады. Ал есептеу техникасы ондай сөздіктер мен грамматикаларды қалаған тәртіпте, бағдарламаға сай өз жадында орналастыра алады.

1946 – 1947 жылдары ЭЕМ арқылы машиналық аударма жасау идеясы Принстонның ғылыми-зерттеу институтында арнайы мәселе ретінде талқыланады (У.Уивер, А.Л.Бут және т.б.). Ең алғаш бұл идея дешифровка ісінің америкалық маманы Уоррен Уивердің меморандумында бірізді шешімге келтірілді. Ал 1949 жылдың 15 шілдесінде атақты ғалым өз меморандумында көтерілген мәселелерді тіл және есептеу техникасы мен дешифровка салаларындағы екі жүзден астам мамандарға жария етті.

У.Уивер машиналық аударманың принциптік мүмкіндігін мүлде бөлек тілдердің ұқсас қасиеттерімен және әр тілдің өзіне ғана тән формалды сипатта болуымен байланыстырады. Қорыта

айтқанда, У.Уивердің пікірі бойынша, қытайша жазылған кітапты қытайша шифрланған (кодталған) ағылшынша немесе басқа ұлттық тілдердегі кітап деп түсінген жөн (*Машинный перевод*, 1957, 33-45).

Сонымен машиналық аударманың өзінше бөлек ғылыми бағыт ретіндегі даму кезеңі Уивердің меморандумынан бастама алады. Бұл бағытты, шартты түрде, негізгі үш кезеңге бөліп қарастыруға болады (*Hutchins*, 1986; *Кулагина*, 1979; *Марчук*, 1983, 91; *Апресян және т.б.*, 1989, 4-5). Мысалы, ол хронологиялық жағынан былайша айырым табады:

1-кезең: 1949 жылдан бастап 60-жылдардың ортасына дейін;

2-кезең: 60-жылдардың ортасынан 70-жылдардың аяғына дейін;

3-кезең: 80-жылдардан қазіргі кезеңге дейін.

Мұндай үш кезеңге бөлуге негіз болған ғылым мен техниканың дамуындағы қол жеткен нәтижелер екені белгілі. Міне, солардың ішінен лингвистикалық, математикалық, техникалық машиналық аударма жүйесін жабдықтау мәселелері айрықша орын алады.

У.Уивердің меморандумымен жарияланған бірінші кезең 1949 жылы Вашингтон университетінде құрылған ғылыми орталықтың зерттеулерінен (ағылшынша-орысша машиналық аударма жүйесі) бастама алады. 1950 жылы А.Д.Буттың ғылыми жетекшілік етуімен французша-ағылшынша аударма орталығы құрылады. Кеңес Одағында 1955 жылы Мәскеудегі КСРО ҒА-ның В.А.Стеклов атындағы Математикалық институтында О.С.Кулагинаның жетекшілігімен ең бірінші машиналық аударма орталығы ашылады.

Осы кезеңде ең алғашқы зерттеушілерді үміттендіретіндей нәтижелер алынып, алғашқы машиналық тәжірибелер жүргізілген болатын. Мысалы, 1954 жылы АҚШ-та Джорджтаун университетінде проф. Леон Е.Достердің жетекшілігімен GAT жүйесі негізінде құрастырылған орыс тілінен ағылшын тіліне аударатын машиналық аударма жүйесі жұртшылық назарына ұсынылды. Бұл бір тілден екінші тілге ЭЕМ (IBM-701) арқылы орындалған ең алғашқы аударма тәжірибесі болып саналады. Ғылымда бұл тәжірибе Джорджтаундық эксперимент деген

атпен тарихта қалды (*Машиный перевод, 1957, 171-б.*). Аталған экспериментте «тікелей» машиналық аударма жағдаяты (бір тілдік жұпқа қатысты жүйелер) тексерістен өткен болатын. Жүйе сөздігінің көлемі бар болғаны 250 сөзді ғана қамтыса, ал оның грамматикасы тек алты синтаксистік ережеге ғана негізделеді (*Machine Translation, 1987, p. 22-23*). Кеңес Одағында француз тілінен орысшаға аударатын машиналық аударманың бірінші эксперименттік жүйесі О.С.Кулагина мен И.А.Мельчуктың авторлығымен 1955–1956 жылдары құрылғаны осы сала мамандарына белгілі (*Кулагина, Мельчук, 1956*). Бұл жүйедегі негіз сөздер сөздігі 1236 сөз, ал сөз айналым сөздігі 250 бірлікті қамтиды. Аударма бірлігі ретінде толық сөйлем түрі есептеледі. Морфологиялық талдау алгоритмі жалғаулардың кестесімен әрекет ету арқылы аударылатын фразаның (сөйлемнің) сөздеріне қажетті деген грамматикалық ақпаратты тіркейді. Бұл жүйенің артықшылығы – омонимдік сөздерді ажырата алатын топтамасы (блогі) болуында. Орыс тіліндегі фразаны синтездеу топтамасында сөздердің орналасу тәртібінің дұрыстығын тексеру мүмкіндігі алдын ала қарастырылған. Сонымен бірге, Т.Н.Молошнаяның жетекшілігімен ағылшынша-орысша аударма жасау жүйесінің алғашқы жобасы ұсынылды (*Молошная, 1960*).

Жаңа бағыттың қарқынды дамуы көптеген мәртебелі ғылыми конференцияларда, симпозиумдарда, жарияланымдарда көрініс тапты және машиналық аударма проблемасымен айналысатын бірнеше ірі орталықтар, топтар мен зертханалар ашылуына себепші болды. Өткен ғасырдың 60-жылдарының басында олардың әр елдегі сандық көрсеткіші мынадай: ФРГ – 22, АҚШ – 21, КСРО – 10, КНР мен Жапония – 6-дан, Франция – 5, Чехославакия – 4, Ұлыбритания мен Венгрия – 3-тен, ГДР мен Италия – 2-ден. Осындай топтар Бельгияда, Болгарияда, Израильде, Канадада, Люксембургте және т.б. елдерде де пайда бола бастады (*Автоматический перевод, 1967, 509*).

Машиналық аударма жұмысының бастамасы жалпы алғанда өрлеу сипатында болды – тілшілер зерттеудің жаңа бағытын ашты, ал программистер күн тәртібіндегі күрделі және аса қызықты мәселелердің шешімін табуға қызыға кірісті. Бұл кезеңде тіл мен аударма процесіне қатысты барлық мәліметтерге

қол жеткендей сезінілді. Сондықтан егер оларды формалды пішінге келтіру ісі жүзеге асырылса, машиналық аударманың әмбебаптық жүйесі құрылып, алға қойылған негізгі мақсатқа тез арада жететіндей көрінген болатын. Бірақ бастапқы және кейінгі тәжірибелердің нәтижелері бұл мәселенің шешімі ойлағандай оңай табылмайтынын байқатты.

Солай бола тұра, атқарылған жұмыстар тек машиналық аударма аясындағы зерттеулерге ғана емес, сонымен бірге ғылыми ізденушілерді тілдің теориялық және құрылымдық жағына да көңіл аударуға ынталандырды.

Қорыта айтқанда, автоматты аударманың бірінші кезеңі белгілі дәрежеде айтарлықтай нәтижелі болды. Себебі, аталған кезеңде көптеген іргелі мәселелердің бастамасы анықталып, зерттеудің болашағы сараланып, жолдары айқындалды. Мысалы, автоматты сөздіктерді құрастыру, аралық-тілді құрастыру, сөздердің омонимдік сипаты мен синтаксистік құрылымы, яғни сөздердің сөйлемдегі тіркесімі, фразеологизмдердің жасалу жолдары және көптеген басқа да мәселелердің сыр-сипаттары анықталды.

Осы мезгілдерде, бүгінгі таңда да аса маңызды деп саналатын, синтаксистік талдаудың алгоритмінің алғашқы нұсқалары зерттелді: синтаксистік тәуелділік пен тікелей құрастырушыларды, болжамдық талдауды, көпвариантты синтаксистік талдауды, «фулькрумдер» әдісімен талдауды, конфигурациялық талдауды және т.б. әдіс-тәсілдерді ескере отырып, талдау жүргізу жақтары қарастырылды.

Терминтану саласындағы бағыныңқылықтың моделі мен тікелей құрастырушылардың моделін талдаудан толық сөйлемнің құрылымдық байласымын анықтауға болатындығы байқалды. Мұндай құрылымның көпмағыналылығының басын ашып алу үшін әртүрлі «сүзбелердің» әдістері қолданылады (қара: ЛМУ-дың математикалық лингвистика зертханасындағы С.Я.Фитиалов пен Г.С.Цейтиннің жетекшілік етулерімен жүргізілген жұмыстар; GETA атты Гренобльдық топ – Centre d'études pour la traduction automatique, жетекшісі В.Вокюа). «Күшті» сүзбе ретінде Д.Е.Хейстің байланыс «күші» ұғымы ұсынылды (The Rand Corporation Automatic Language Data Processing Group. Santa Monica (Calif.)). Байланыс күшінің

жоғары дәрежесіне қарай талдау варианттарының артықтық деңгейі айырым тапты (*Автомат. перевод, 1971. 41-б.*).

Болжамдық талдау сөздік мақаланың айрықша құрылымына негізделеді (Национальное бюро стандартов США, И.Родес; Вычислительная лаборатория Гарвардского университета (Масс.), рук. А.Эттингер). Мұнда әрбір сөзге оның барлық байланыс мүмкіндігі жайлы ақпарат тіркеледі. Талдау тәртібі (процедурасы) мәтінді солдан оңға қарай тексеріп шығу тәсілі бойынша жүргізіледі. Осының нәтижесінде әрбір кезекті сөздің «өлшемінің» алдыңғы сөз бойынша жасалған болжамға сай келетіндігі тексеріледі. Егер ол жасалған болжамға сай келмесе, ондай сәйкестікті іздестіру сол жаққа қарай ауысады. И.Родестің ғылыми тобында бірварианттық талдау қарастырылады, яғни талдауға түскен сөздің сол жағында орналасқан бірінші сөз сәйкестік «иесі» деп саналып, кез келген сөйлемге тек бір ғана синтаксистік құрылым сәйкес қойылады.

Сөйлемнің синтаксистік құрылымының бірнеше варианттарын есепке алуға мүмкіндік туғызатын көпварианттық талдауды бір қадам алға басу деп санауға болады (С.Куно, А.Эттингер, Гарвард университетінің Есептеу зертханасы). Көпварианттылық талдауда сөздің синтаксистік потенциалды мүмкіндіктерінің молдығының арқасында синтаксистік байланыста тұратын сөздерге болжам жасауға болатындығы негізге алынады. Көпварианттылық талдаудың алғашқы грамматикасы 2100 ереже мен 82 синтаксистік болжамды қамтыды. Сөйлем құрылымының диаграммасы тікелей құрастырушылар моделі негізінде құрылды. Осының нәтижесінде мүмкін болатын барлық варианттар іріктеліп, олардың көпшілігі әртүрлі сүзбелер көмегімен сан жағынан шағындалды. Болжамдық талдау алгоритмінде көпвариантты фразаның синтаксистік құрылымына тән екімәнділік ақиқаттығынан гөрі, оны таңдалып алынған варианттарды іріктеу әрекетінің формалды сипатының салдары деуге болады. Міне, бұл жүйелердің осалдығы да осында. Дегенмен, синтаксистік болжам жайлы идеяның өзі де құнды деуге болады. Бұл ой-пікірге қазіргі кездегі кейбір грамматикалық болжамдарға сүйенетін автоматты талдау жүйелері негізделуде (*Цейтин және т.б., 1982; Откупщикова, 1982*).

Атап айтуға тұратын жайт — бірварианттық пен көпварианттық талдауларды бір-біріне қарсы қою машиналық аударманың қалыптасуының бірінші кезеңі үшін ерекше маңызды болуы. Бірварианттық талдауда кездейсоқтыққа жақын дәрежеде тандалып алынған сөйлем құрылымы дұрыс болмай шығуы мүмкін, ал талдау жұмысын басынан қайталау мүмкіндігі бұл жүйелерде алдын ала қарастырылмаған. Көпварианттық талдауда біркелкіліктің бұзылу қаупі бар: құрастырылған құрылымдардың саны вариантты дұрыс таңдау мүмкіндігінен де айтарлықтай үлкен болуы мүмкін.

Сонымен бірге жақсы аудармашы (немесе автоматты аударма жүйесі) мәтін мазмұнын түсіну үшін аударуға тиісті фразаның дұрыс түрдегі бір ғана вариантын ұсынуы керек. Ал фразалардың көпмағыналылығы мәнмәтін бойынша өз шешімін табатыны белгілі, сондықтан сөйлемдердің нақты мағыналарын ажырату мәселесі мәтінде баяндалатын ақиқат жағдаятқа қатысты толық мәтіннің синтаксистік және семантикалық заңдылықтарына тікелей байланысты болып келеді. Мұндай семантикалық мәліметтер машиналық аударманың бірінші және екінші кезеңдерінде үзінділер түрінде ғана көрініс тауып, толығымен қолданыла қоймады. Себебі, машиналық аударма жүйесін тиісті лингвистикалық мәліметтермен жабдықтау ісі әлі де болса толық зерттелмеген болатын.

Конфигурациялық талдау әдісі аударылуға тиісті мәтінді (жүйеге енетін мәтін) сол тілдің алдын ала іріктеліп алынған синтаксистік конструкциялар жиынымен салыстыру тәсіліне негізделеді (*Каф. матем. Лингв. ЛГУ; Белопольская, Ковригин, 1963; Воронин, 1961. Каф. Струк. и прикл. Лингв. МГУ; Дрейзин, Рашитов, 1961. Группа авт. перевода ИЯ АН СССР; Молошная, 1960 және т.б.*). Егер аударма жүйесіне енетін мәтіннің анықталған конфигурациясы жиынның синтаксистік конфигурацияларының біреуімен сәйкес келсе, онда мәтіндік конфигурация танылды деп саналады және ол әрі қарайғы талдау әрекетінде ықшамдалған түрінде көрініс табады. Синтездеу әрекетінде анализдеу кезінде анықталған конфигурация аударуға тиісті тілдің оған сәйкес келетін конфигурациясымен салыстырылады.

«Фулькрумдер» әдісі бойынша жүргізілетін талдауда мәтіндегі сөйлем төменгі сатыдағы бөліктерге ажыратылады. Осының негізінде кез келген фразаны циклдық әдіс бойынша бірнеше рет қарастыра отырып, іріленген фрагменттерге біріктіруге болады. Әрбір кезекті циклдық қарастыру кезінде талдаудың тірек нүктелері (*fulcrum*) – фулькрумдер анықталады. Анықталған фулькрумдер синтаксистік конструкциялардағы негізгі және «басқарушы» сөздер тобына жатады. Олардың құрылымы мен шекарасын анықтау циклдық қайталаулардың мақсатына енеді (*группа автомат. Перев. фирмы Bunker Ramo-Woldridge Inc, Canoga Park, Calif., рук. П.Гарвин*).

Тізбекті талдау әдісі арқылы фразаның синтаксистік құрылымын тану үшін берілген сөйлем көрінісіндегі синтаксистік категориялар тізбегін элементарлық тізбектерге бөлу қажет болады (*Харрис, Н.Сейгер*).

Корреляциялық талдау кезінде әрбір сөзге индекстер жиынтығы тіркеліп жазылады. Олардың әрбіреуі сол сөздің осыған ұқсас индексі бар басқа сөздермен өзара байланыстылық (корреляция) мүмкіндігін бейнелейді. Корреляцияның (өзара қатынастылық) жоғары деңгейі лингвистикалық шектеулер арқылы төмендетіледі. Мысалы, сөз, семантика және т.б. сатыларға төмендейді.

Осымен бірге зерттеушілер машиналық аударма мәселесі үшін толық мәтінді талдау мәселесін қарастыра бастады. Соның ішінен толық мәтіндердің антецедент – зат есімдермен арақатынасын анықтайтын ережелердің қарастырыла бастағанын атап кетуге болады (антецедент – анафоралық қатынас. Мәселен, есімдіктен бұрын тұратын есім сөз, ал есімдік осы есімге нұсқайды) (*Шумилина, 1961; Автомат. перевод, 1967, с. 235 к сл.*).

Қазіргі кездегі жоспарланып жатқан машиналық аударма жүйесі екі топқа бөлінеді: бір тілден екінші тілге аударатын – бинарлы және бірнеше тілдерден бір тілге немесе бірнеше тілдерге аударуға арналған – көптік. Машиналық аударма жүйелерінің бірінші түрінің әр уақытта артықтық сипаты адам баласының аудару әрекетін модельдеудің және әрбір тілдер жұбы деректерінің бинарлық сәйкестіктерін анықтаудың теориялық және практикалық мүмкіндіктерінің табиғилық жағдаяты-

нан туындайды. Машиналық аударманың бинарлық жүйелерінің кемшілігі — аударуға тиісті «кіретін» тілге талдау жүргізу кезінде олардың қасиеттеріне қатаң түрде тәуелді болуы және әрі қарайғы аударма алгоритмін құрастыруда ол тәуелділіктің күрделене түсетіндігі.

Көптік аударма жүйесі аталған қатаң тәуелділікте болуына қарсы мүмкіндіктер тудырып, талдау мен жинақтаудың тәуелсіз алгоритмдерін құрастыруға жол ашты. Бинарлық аудармалардың қосындысы ретінде түсіндірілетін көптік аудармадағы тіл жұптарының саны $n(n-1)$ формуласы арқылы анықталады. Бұл өрнектегі n — аудару процесіне қатысатын тілдер жиынының саны.

Көптік жүйлерді зерттеуші ғалымдар талдау мен жинақтаудың баламасы (инвариантты) ретінде **аралық-тіл** идеясын (интерлингва) ұсынды (*Машинный перевод, 1957*).

Көптік аударма жасаудағы «аралық-тіл» идеясын іске асыруда ең алғашқы елеулі орынға ие болатын ғалымдардың ішінен Н.Д.Андреевтің аты бөліп айтуға тұрарлық (*Андреев, 1957*). Н.Д.Андреев жетекшілік еткен зертхана (ЭЛМП) қызметкерлерінің осы бағыттағы көптеген ғылыми еңбектерінің жарияланғаны осы саладағы мамандарға мәлім. Осы кезде Кеңес Одағы мен шетелдердің ғылыми орталықтарымен басқа да ғылыми топтар арасында *аралық-тіл* мәселесіне қатысты ой-пікірлер қызу түрде талқыға түсе бастады.

Н.Д.Андреевтің концепциясы бойынша, *аралық-тіл* ұғымының негізінде табиғи тілдердің табиғатында конгруэнттік сипаттың болмайтындығы жайлы түсінік жатыр. Яғни аралық-тілдің жасандылығының заңды екендігі, сол себепті одан конгруэнтті емес элементтердің (артикльдер және т.б.) алынып тасталуы мен көптеген тілдер үшін конгруэнтті элементтердің (етістіктің үш түрлі шағы, көптік жалғау және т.б.) сақталып қалуы қажеттігі туындайды. Сондықтан аталған мәселеге арналған алғашқы зерттеулерге *аралық-тіл* өзінің құрылымы жағынан табиғи тілге жақын тұрып, соның терминдерімен баяндалады.

Әр тілдің бірліктері бойынша грамматикалық мағыналарының, грамматикалық категорияларының және бәріне ортақ болып келетін семантикалық мағыналарының салыстырмалы

сипатының қорытындылаушы қасиеттерінің белгілі болуына қарамай, семантикалық теорияның қолданбалы тұрғыда жеткіліксіз зерттелуі тілшілерге дағдылы тілдік көзқарастан ауытқуына кедергі жасады (Мельчук, 1959; Автомат. пер., 1967). Н.Д.Андреев жетекшілік еткен машиналық аударманың тәжірибелік зертханасында (ЭЛИМП) ЭЕМ-ның көмегімен 18 шет тілінен орыс тіліне аударатын және аралық-тіл идеясын жария ететін шағын дәрежедегі тәжірибе жүргізілді.

Аралық-тіл идеясын өмірге әкелу мақсатындағы сан алуан әрекеттерге қарамай, көптеген тілдерден бір тілге аудару тәжірибесі машиналық аударма тарихында айтарлықтай өз орнын ала алмады деуге болады. Теориялық тұрғыдан қарастырғанда оның себебі аралық-тілдің грамматика мен лексика терминдерінде сипатталуында деп түсіндіріледі. Шындығында, ол тек қана семантикалық тұрғыда, семантикалық барабарлық түсінігінде көрініс табуы қажет болатын.

Сыншылар аралық-тіл арқылы аударма жасаудың төмен дәрежедегі барабарлығын айта келе, оның аударуға тиісті мәтіннің маңызды сипаттамаларының назардан тыс қалып қоятындығын айтады: мәтін авторының коммуникативтік және прагматикалық ұстанымы, талдаудың актуалдығы және т.б.

Машиналық аударма жүйесіндегі аралық-тілдің жүзеге аспай қалуын лингвистикалық зерттеулердің практикалық тұрғыдағы көзқараспен түсіндіретін болсақ, ол өзгеріс енгізу кезеңдерінің әдеттен тыс күрделілігінен, ал техникалық жағынан алғанда ЭЕМ-ның алғашқы буындарының жад көлемі мен әрекет жылдамдықтары көрсеткіштерінің төменгі дәрежелігі деуге болады. Мұндай кедергілер бинарлы аударманың экспериментті жүйесін құру жағдайында да теріс әсерін тигізді.

Қазіргі кезде аталған кедергілер әртүрлі аралық-тілді іске қосу кезінде өз шешімін табуда. Аралық-тіл мәселесіне қызығушылықтың екінші айналымы — жасанды интеллект жүйесін құрастыруға байланысты. Себебі, тек осы идея бойынша ғана қазіргі кездегі кибернетика саласында зерттеліп жатқан мәтіннің мазмұнын түсінетін интеллектуалды жүйелерге жол ашылды деуге болады. Бірақ бүгінгі таңның өзінде-ақ аралық-тіл арқылы машиналық аударма жұмысын жүзеге асыру

кез келген аударма жүйесін құрастырушының арманы, жоғарғы мұраты деуге болады. Солай бола тұра, аралық-тіл мәселесін зерттеу семантикалық теорияның, табиғи тілдің фразалары мен мәтіндерінің семантикалық көрінісін анықтау жағдаяттарының даму процесіне зор ықпалын тигізді.

Машиналық аударманың дамуына **аралық-тіл концепциясының** ешбір күмәнсіз үлесі – аударылған және аударылатын тілдердің қасиеттеріне ғана негізделетін анализ бен синтездің бір-біріне тәуелсіз екендігін мойындау; аударылатын сөйлемнің мән-мағынасының үйлесімділігі мен әмбебаптық берілісін зерттеу қажеттігін мойындау болып табылады.

Машиналық аударма жүйелерін құрастыру аясындағы жүргізіліп жатқан жұмыстардың белсенділігіне қарамай, бірінші кезеңде зерттеушілердің ойлағанындай практикалық нәтиже алынбады. Аударманың тұтынушы сұранысын қанағаттандыра алатын өндірістік жүйесін құру жұмысының нәтижесіз болуының үлкен зардабы байқалды. Мысалы, зерттеушілердің бір бөлігі машиналық аударма мәселесін шешуден бас тартты. Осы бағытта жұмыс істейтін бірнеше топтың зерттеу әрекеттерінің тоқырап қалуына себепші болған АҚШ-тың Ұлттық Ғылым академиясының арнайы комиссиясының әуе соғыс күштерінің машиналық аударма жайлы жасаған талдау қорытындысын бұрмалап түсінудің (экстраполяция) де өз рөлі болды. Осының зардабынан көптеген ғылыми ұжымдар ұйымдастыру мен қаржылық қиындықтарға әкелді (*Кулагина, 1979, 14*).

Мұндай жағдайға душар болу кейбір ғалымдардың скепсистік көзқарасы да, машиналық аударма идеясының жүзеге асу мүмкіндігіне сенімсіздік білдіруі де, әсіресе, ғылыми-техникалық мәтіндерді аудару мүмкіндігіне күмән туғызуы да себепші болды. Ал көркем әдебиет мәтіндері, оның ішінде, әсіресе поэзия мәтініне байланысты машиналық аударма жасауға ең басынан-ақ ғалымдар қарсылық пікір білдірген болатын. Оған шығарманың көркемдік, бейнелік жағын ұстаным етудегі кейбір тілдік заңдылыққа қайшы келу жақтары негізгі себеп болды.

Осындай жағдайға байланысты И.Бар-Хиллелдің көзқарасына тоқтала кетуге болады. Ғалым жоғары сапалы толық

автоматтандырылған машиналық аударма мүмкін емес және оны утопия деп санаған. И.Бар-Хиллелдің пікірінше, машиналық аударма кезіндегі негізгі кедергі сөздердің көпмағыналылығы және тіл табиғатында ондай екімәнділікті ажырата алатын формалды көрсеткіштердің болмауы. Бұл қиындықтардың шешімі тек әлем туралы білімде жатыр деп тұжырым жасайды (*Бар-Хиллел, 1962*).

Жоғарыда айтылған сенімсіздік пікірлер әлем бойынша машиналық аударма мәселесімен айналысатын топтар санын азайтуға әкеліп соқтырды, екінші кезеңнің аяғында 100 топтан 38 топқа дейін азайғанын айтуға болады (*Кулагина, 1979, 16-18*).

Бірақ машиналық аударманың болашағына сенімді, дәйекті көзқарастағы ғалымдар қажымай, шаршамай-шалдықпай жоспарлаған ғылыми ізденістерін жалғастыра бергені сала мамандарына мәлім. Енді машиналық аударманың екінші кезеңі басталды. Машиналық аударма мамандары зерттеудің теориялық жағын тереңдете түсіп, формалданған лингвистикалық модельдер жасауға, жүйелі және функционалды грамматика жазу мәселесіне кірісті. Бұл кезеңде семантиканы модельдеу жұмысына арнайы назар аударылды. Бірінші кезеңнің нәтижелері сараланып, болашақтағы прагматикаға бағышталған жүйе құрудың мүмкіндігі анықталды.

Бірінші кезеңде зерттеле бастаған формалды грамматиканың модельдері: теоретика-көптік модельдер, бағыныңқылықты грамматикалардың модельдері, тікелей құрастырушылар, туындату грамматикасы мен семантикалық модельдер және т.б. Екінші кезеңде машиналық аударма мәселесінің зерттелуі өз жалғасын тауып, оның интерпретациясына, яғни түсініктеме берілу жағына көбірек көңіл бөлінді, машиналық аударма мәселесі жеке тілдердің грамматикалық құрылымының аздықөпті толық түрдегі сипаттамаларында да қолданыс тапты. Сонымен бірге, олар автоматты талдаулардың іргелі алгоритмдері бойынша да сыннан өтті деуге болады.

Екінші кезеңге қатысты мынадай жайттарды атап кетейік: семантикалық талдау принциптерінің бұрынғыдан да қарқынды түрде зерттелуі және жоғары сапалы аударма жасауға қатысты талдаулар жасау жүйелерінде семантиканың қатысынсыз мәселенің шешілмейтіндігін толық мойындау.

Соңғы айтылған жайтқа байланысты, яғни семантикаға байланысты: біріншіден, кейбір машиналық аударма жүйелерінде талдаудың айрықша семантикалық деңгейін (деңгейлерін) енгізу арқылы (мысалы, ФРАП), ал екіншілерінде – алдын ала белгілі синтаксистік қатынастар типтерінің кең түрдегі синтаксистік сипаттағы тізбесі арқылы мәселені шешу (мысалы, «мән-мағына – мәтін» моделінде), үшіншісінде – аралық-тіл типіндегі мән-мағына инвариантын ұсыну арқылы және осыған әкелетін семантикалық ережелерді қолдану арқылы және т.б. Машиналық аударма жүйесін құруға байланысты жұмыстардың екінші кезеңде жалғасуымен бірге, ол кезең аударманы автоматтандыруды тәжірибелік сынан өткізу және оны өндіріске енгізу әрекеттерімен қоса сипатталады.

Сөздік пен базалық алгоритмдердің көлемдерін ұлғайту арқылы Джорджтаундық жүйе (GAT) және олардың жекелік түрлері даму үстінде болды (мысалы, SYSTRAN жүйесі). 1976 ж. бұл жүйенің ағылшынша-французша версиясы Еуропалық Экономикалық бірлестік қоғамына тапсырылды. Аталған жүйеде негізі синтаксистік байланысты тағайындайтын талдау тәсілі ескерілген болатын (Марчук, 1985). Анализ бен синтездің бір-біріне барынша тәуелсіз болуынан оларды басқа салалас тілдік жүйелерге де қолдану мүмкіндігі туындады.

Өткен ғасырдың 70-жылдарынан бері зерттеліп келе жатқан Гренобльдық орысша-французша аударма жүйесінің (GETA) жетілдірілген түрі логикалық пішіндегі терең синтаксистік құрылымдарды алуды көздейтіні мәлім. Бұл жүйе бойынша семантикалық, синтаксистік, морфологиялық және сөздік деректері арқылы бірнеше өндеулер (трансформаций) жүргізе отырып, «шығыс» тілдің (выходной язык) үстірт тұрпатының (поверхностная форма) көрінісін анықтауға мүмкіндік туады. Көпдеңгейлік талдау өндеудің 11 топтамасының реттік тәртібі түрінде құрылады. Мысалы, аналитикалық тұлғаны синтетикалыққа келтіру, фразаны сегменттеу, есім топтарын талдау және т.б. Мұндай әрекеттер антецеденттерді іздестіру мен қалпына келтірілген түйісу нүктелерінің синтаксистік және логикалық функцияларды анықтауға дейін өз жалғасын тапты. Аталған жүйе ARIANE атты дамыған программалық (бағдарламалық) жабдықталуға ие болды. Мысалы, ол жүйе мен тұтынушы

арасындағы әрекетті деректер базасы арқылы алдын ала қарастырумен қатар, процесті басқару мен мәтінді өңдеуде де аса дамыған құралдар қолданыс тапты.

ТАУМ жүйесінде (Монреаль университеті, Канада, жетекшісі — Р.Китредж) ағылшынша-французша аударма үстірттік (поверхностный) синтез деңгейінде жүзеге асады (*Синтаксический компонент ...*, 1981). Аталған жүйе лексикалық және құрылымдық жағынан өңдеу міндетін атқаратын трансфер топтамасына (блогына) ие. Оған лексемаларды аудару, француз тіліндегі фразаны синтездеуді жеңілдету үшін құрастырушылар таралымының (дерево) кейбір бөліктерін өңдеу, етістіктің шақтық түрін және меңгеру моделін (модели управления) аудару мәселелері қамтылды. Программаны (бағдарламаны) жазу тілі ретінде А.Кольмер құрастырған Q (Quebec) тілі алынған (*Синтаксический компонент...*, 1981, с. 92). Бұл жүйедегі талдағыштың комбинаторлығы мен икемділігі (талдаудың алғашқы болжамын аяғына дейін сақтау, ережелердің қолдану тәсілдерінің ретін белгілеу және т.б.) аударманың үстірттік синтаксиске бағышталуына мүмкіндік тудырады.

ТАУМ жүйесі 1968 жыл мен 1980 жыл аралығында белсенді түрде дамып, нәтижесінде 1977 жылдан бастап аталған жүйе жыл сайын 5 млн. сөз көлеміндегі мәтіндерді аударуды жүзеге асырып, өндірістік жүйе дәрежесіне дейін көтерілді (*Machine Translation* 1987, p. 12).

Чехославакиядағы Карлов университетінің алгебралық және есептеу лингвистикасы тобы П.Сгаллдың ғылыми жетекшілігімен жай сөйлемдерді ағылшын тілінен чех тіліне шектеулі сөздік бойынша аударма жасайтын машиналық аударма жүйесін құрастырды (*Синтаксический компонент...*, 1981, с. 84). Бұл жүйеде талдау процесі негізгі төрт кезеңнен тұрады: морфологияға дейінгі талдау, морфологиялық талдау, синтаксистік талдау, аудару. Синтаксистік талдау синтаксистік ережелердің әрекет ету аясының кеңейтілген типі ретінде құрылған: алдымен шағын мәнмәтіннің бірліктері талданады және әрі қарай оның көлемі біртіндеп ұлғайып, сөйлем шеңберінен сырт шығатын мәнмәтін көлеміне дейін өседі. Сондықтан мұны әдеттегі «төменнен жоғары қарай» талдау жағдайына жатқызуға болады. Осыған байланысты синтаксистік

талдаудың өзі де бірнеше кезеңге бөлінеді. Аталған жүйеге жүйесіз (нерегулярный) синтаксисті, мағынаны дұрыс ашуға қажетті сөйлемде жоқ элементтерді толықтыру ережелерін талдау үшін «авариялық» деп аталатын грамматика енгізілген. Синтаксистік қатынастарға семантикалық түсініктеме беру тәсілі кең түрде қолданыс тапқан.

Шамамен алғанда, 1970 жылдан бастап, SUSY жүйесін құрастыру жұмысы басталады (*Синтаксический компонент...*, 1981, с. 90). Аударма жасау процесі деректерді алдын ала дайындаудан, талдаудан, өңдеуден және синтездеуден тұрады. Аударма алгоритмін жасау әр тілдің грамматикасына тәуелсіз сипаттау принципіне негізделеді, сондықтан жүйеге жаңа тілді қосу әртүрлі жаңа бағдарламаларды құрастыру әрекетін сан жағынан шектейді. Неміс, орыс, ағылшын және француз тілдерін талдау, ал сонымен бірге неміс тіліне қатысты синтездеу жұмыстары жүйелі түрде жүргізілді. Туындау таралымы (дерево) түріндегі «базистік құрылымды» — талдау жұмысының нәтижесі деуге болады. Сөздіктер жүйесі (талдау сөздіктері, семантикалық сөздіктер және аударма сөздіктері) әсіресе талдау процесіндегі семантикалық ақпарат үшін аса маңызды болып келетін кез келген ақпаратты пайдалануға мүмкіндік туғызады. Талдау принципі өзара қисынды түрде құрастырылған: талдаудың бірінші бөлігі «жоғарыдан төменге қарай» өңдеуден басталады, екінші бөлігі «төменнен жоғарыға қарай» өңдеуден басталады, ал үшінші бөлікте алдыңғы екі нәтиже бір-біріне тіркесіп орындалады («жапсырылады»). Талдау соңында бұрыс синтаксистік құрылымды өңдеу шаралары қарастырылған.

Машиналық аудармамен айналысатын ресейлік топтардың сан жағынан аз болуына қарамай, сөз болып отырған екінші кезеңде олардың зерттеу жұмыстарын аса нәтижелі деп сипаттауға болады:

а) лингвистикалық және математикалық жабдықтау тұрғысынан алғанда, бастапқы машиналық аударма жүйелерінің жақсартылған түрлерінің құрастырылуы;

ә) жаңаша типті жүйелердің құрастырылуы (Бүкілодақтық аударма орталығы, А.И.Герцен атындағы ЛГПИ-дегі «Статистика речи» тобы);

б) автоматтанған талдау жүргізетін жүйелердің қарқынды түрде зерттелуі;

в) машиналық аударма жүйелерін әртүрлі мәндік дәрежеге қатысты тәжірибелік сынау деңгейіне дейін жеткізілуі және бұрын-соңды құрастырылған жүйелердің жетістіктері мен кемшіліктері бірдей дәрежеде сыннан өту нәтижелері осыны байқатады.

Аталған кезеңде КСРО ҒА-ның Қолданбалы математика институтында французша-орысша машиналық аударма жүйесінің жаңаша түрі (жаңаша версиясы) құрастырылды. Француз тіліндегі математикалық мәтінді орысшаға аударатын бұл жүйе 1967–1976 жылдар аралығында О.С.Кулагинаның ғылыми жетекшілігі бойынша зерттеу жұмыстарын жалғастырды (*Кулагина, 1979*). Жүйеде аударма әрекетіне қатысты өңдеу жұмыстары былайша іске асты: жүйеге «кіретін» ақпарат французша фразаның үстінгі синтаксистік құрылымы (фраза бойынша аудару), ал жүйеден «шығатын» ақпарат – мән-мағыналық сәйкестікте тұратын орысша фразаның үстінгі (поверхностный) синтаксистік құрылымы. Мұндай тәсіл екі түрлі тілдердің сөздері мен фразалары мағыналарының сәйкестігі мен олардың құрылымдық сәйкестік сипатына негізделеді. Бұлайша талдаудың негізгі мақсаты бағыныңқылықтар таралымы (дерево) түріндегі фразалардың синтаксистік құрылымдарының жиынын құру. Осылайша талдау әдісін жүргізе отырып, синтаксистік байланыстың 52 типі жеке түрде бөлініп алынады.

Көпварианттық талдауда сүзбе-ережелер жиыны қолданылады. Мұндай әдіс арқылы синтаксистік байланыстың күші мен арақашықтығы және байланыстың ортақ немесе ортақ еместігі (совместность/несовместность связей), құрылымдардың жобалауға болатындығы немесе болмайтындығы (проективность/непроективность структур) ескеріледі. Сүзбелерді қолданудың ретіне қарай, олардың иерархиялық айырым-белгілері есепке алынған. Мысалы, алдымен қарапайым сүзбе, сосын барып жалпы, күрделі, ал ең соңында жекеленген жағдайға сай келетін сүзбе түрлері қолданылған. Көпвариантты талдау түрлеріне сүйене отырып, болжамдар жиыны ішінен аталған сүзбелер арқылы нақты бір болжамды құрастыруға

мүмкіндік туындайды (Кулагина, 1979; Синтаксический компонент..., 1981).

Орыс тілін синтаксистік талдау арқылы зерттеу жолында машиналық аударманың алғашқы кезеңінде орын алған конфигурациялық талдау идеясы қолданыс тапты. Осының нәтижесінде сөйлем ішіндегі сөздердің әр алуан синтаксистік қатынас типтерін мағыналық тұрғыда сипаттайтын тікелей доминация қатынасы (отношение непосредственной доминации) деп аталатын 31 қатынасы жеке бөлініп алынды. Мұндай синтаксистік талдаудың негізгі мақсаты аударылатын сөйлемнің сөздері арасындағы синтаксистік байланысты анықтау болып табылады. Ол үшін мұндай байланыстарға қажетті тікелей доминация қатынастарын тіркеп жазу және оның мазмұнына қатысты ерекше сипаттамалар сәйкестендірілуі қажет болады. Автоматты талдау жүйесінің мәтінді аудару, реферат алу, ақпаратты іздестіру сияқты қолданбалы мәнімен бірге, теориялық жағы да аса құнды деуге болады. Себебі, талдау алгоритмі «тілдегі талдаушы бөліктің» әрекет ететін және эксперименталды түрде бақыланатын моделі болып табылады (Автоматический перевод, 1967. 226; Мельчук, 1974).

Кезінде Ленинград университетінің математикалық лингвистика зертханасында (НИИММ) ағылшынша-орысша аударма жүйесі тәжірибелік тексеру дәрежесіне дейін жетті (Лейкина және т.б., 1966). Бұл жүйеде талдау мен жинақтауды бір-біріне тәуелсіз қарастыру мүмкіндігі есепке алынды. Ал өңдеу мен түсініктеме беру әдістерінің міндеттері – бірізділікке келтіру. Талдау әдісі фразалық және көптік сипатта жүргізілді. Бағыныңқылық грамматикасы моделінің шеңберінде тұжырымдалған лингвистикалық мәліметтер, оларды өңдеу шараларынан (алгоритмдерден) жеке қарастырылды. Ал алгоритм жұмысының нәтижелігі синтаксистік талдау әдісін дұрыс жүргізу, жобалау, тарамдалу (древесность), байланыстың міндеттілігі, ортақтығы және т.б. шарттарын қамтамасыз ету арқылы жүзеге асты.

Көп ұзамай-ақ, жоғарыда аталған зертханада байланыстың тұйықталу сипатына айрықша мән беру мен реттілікті есепке алатын орыс тілінің синтаксистік жүйесін құру және оны

тәжірибелік сынақтан өткізу жұмысы жүргізіледі (*Кремнева және т.б., 1976*).

Ю.Д.Апресянның жетекшілігімен ИНФ-ЭЛ атты лингвистикалық жағынан аса күшті жабдықталған французша-орысша аударма жасау жүйесі құрылды (*Синтаксический компонент..., 1981*).

Аударма үстінгі (поверхностный) синтаксистік құрылым деңгейінде жүргізіліп, мына принциптерге негізделді:

1) аударылуға тиісті тілдердің грамматикаларының сипатталуы «мән-мағына-мәтін» («смысл-текст») атты бірыңғай модель негізінде құрылуы;

2) талдау мен жинақтау ережелерінің бір-біріне ауысу мүмкіндігіне және олардағы ақпарат көлемінің жеткілікті болуы;

3) грамматика мен сөздікке қатысты сипаттаулардың алгоритмдер мен программалардан бөлектік сипаты (аудармаға қатысты ережелерден басқа);

4) аударуға қатысты тілдердің сипаттауларының тәуелсіздігі;

5) әртүрлі пәндік салаларға бейімделуінің әмбебаптылығы;

6) лингвистикалық білімді бейнелеу үшін формалды аппаратты (модель және тіл) құру.

Бірінші кезекті ИНФ-ЭЛ жүйесі «ИРИС-50» ЭЕМ-да жүзеге асты.

Мәскеудегі М.Торез атындағы шет тілдер институтында АРМАС атты машиналық аударма жүйесі үшін көпаспектілі ағылшынша-орысша сөздік құрастырылды. Оның мақалалары сөздік бірлігінің әр алуан лингвистикалық сипаттамаларын қамтиды: морфологиялық, синтаксистік, семантикалық, лексикалық қызметі жайлы мәліметтер, меңгеру моделіне қатысты аударма баламалары жайлы және т.б. (*Шаляпина, 1974*).

Бұл сөздік АРАП атты ағылшынша-орысша автоматты аударма жүйесінің лингвистикалық жабдықталуының маңызды бөлігі ретінде түзілді (жетекшісі –З.М.Шаляпина). АРАП жүйесі үшін қабылданған ұстаным ЯРАП атты жапонша-орысша автоматты аударма жүйесі үшін де қолданылды.

Аталған жүйелер үшін талдау жұмысы семантика-синтаксистік деңгейде жүргізілді. Тілдік бірліктердің семантика-

лық тіркесін зерттеу мақсатында арнаулы тіл құрастырылып, барлық негізгі лингвистикалық ақпараттар біртүптік сөздік мақалалары түрінде көрініс тапты. Оларда сөздік бірліктері грамматикалық, лексикалық, тіркесімдік және басқа да әр фразаны аударуға жеткілікті белгілер түрінде көрініс тапты.

Бүкілодақтық аударма орталығы жұмысының екінші кезеңінде бірнеше аударма жүйелері құрастырылды. Олардың ішінен тек екеуін ғана атап өтейік: Н.Н.Леонтьева жетекшілік еткен ФРАП атты французша-орысша автоматты аударма жүйесі (*Леонтьева, 1987*) және Ю.Н.Марчук жетекшілік еткен АМПАР атты ағылшынша-орысша автоматты жүйе (*Марчук, 1983*).

ФРАП жүйесі аудармашының әрекетін модельдеуге арналған аса қызықты талпыныс деуге болады. Аударылатын мәтін тілінің аудармашының ана тіліне қатысты еместігі ескерілді, сондықтан мәтін ішінде аудармашыға белгісіз тілдік элементтердің кездесетіндігі де есепке алынған. Ғылыми тұрғыдағы мұндай зерттеудің мақсаты – таза грамматикалық тәсілдің «күшін» анықтау. Осыған байланысты аударма процесі қажетті фразаны талдау деңгейінің белгілі сатысына сәйкес келетін көптеген топтамаларға (блоктарға) бөлінеді. Ең алдымен формалды-грамматикалық құралдардың барлық мүмкіндіктері пайдаланылады, ал содан кейін ғана талдаудың семантикалық деңгейі іске қосылады. Семантикалық деңгейдің грамматикалық деңгей тәріздес, сан жағынан бірнешеу болуы алдын ала ескерілген. ФРАП жүйесіндегі аударманың концепциясы жүйенің синтаксистік компонентінің «күшін» бағалайды. Бұл жүйенің ерекшелігі – аса дамыған семантикалық аппаратпен әрекет етуі. Бірінші синтаксистік топтаманың жұмысы аяқталғаннан кейін ғана келесі семантикалық аппаратқа бетбұрыс басталады. Жүйенің әр деңгейіндегі ішкі кішігірім жүйелерді тәжірибелік сынақтан өткізу де алдын ала қарастырылған.

Грамматикалық және дамыған лексикалық талдауларға бағышталған АМПАР жүйесі де көпдеңгейлі. Онда сөз айналымдары, бірмағыналы және көпмағыналы сөздер жеке топтамалар арқылы өңделіп, тек сөйлем құрылымдары ғана өзгеріске ұшырайды. Аталған жүйе іс жүзінде қолданыс табуда.

Белгілі ғалым, профессор Р.Г.Пиотровский жетекшілік ететін Бүкілодақтық «Статистика речи» тобы 1970 жылдан бастап машиналық аударма жүйесі бойынша зерттеулер жүргізді. Бұл жүйе қазіргі уақытта да актуалды болып саналатын итеративті-стратификациялық тәсілге негізделеді. Мұндай әдіс бойынша аударуға тиісті мәтінді қайшылықсыз сипаттау үшін ол кезеңдік сипатта қарастырылуы қажет.

Ленинградтағы А.И.Герцен атындағы Педагогикалық институт-тың инженерлік лингвистика зертханасында 1975 жылдан бастап зерттеле бастаған СИЛОД атты аударма жүйесі көптілді жүйе ретінде өз міндетін атқара бастады. Оның сөздіктермен және морфологиялық тұрғыда жабдыкталуы аса күшті және құрамында трансфер қолданыс табатын грамматикасы бар жүйе болып саналады.

Осы кезеңдерде ағылшынша-орысша және французша-орысша машиналық аударманың түрлері құрастырылып, тәжірибелік сыннан өтеді. Осы әрекеттердің нәтижесінде жекеленген синтаксистік конструкциялардың құрылымы мен мағынасын тануға мүмкіндік тудыратын талдаудың фреймдік және тезаурустық негіздері қаланады (Пиотровский, 1979; Чижасковский, Беляева, 1983). Аталған жүйеде машиналық аударманың алғашқы кезеңінде ұсынылған «фулькрумдар» әдісінің дамыған түрі ретінде саналатын машиналық аударма алгоритмдерін деңгейлік тұрғыда құрастыру ескерілген.

Машиналық аударма жүйелерін құрастырудың екінші кезеңін қысқаша қарастыра отырып, олардың негізі сипаттамалары мен жалпы ерекшеліктерін былайша атап кетуге болады (Кулагина, 1979; Hutchins, 1986):

1) көпдеңгейлі талдау мен синтаксистік деңгейдің толық дәрежеде көрініс табуы;

2) талдау нәтижелерінің көпварианттылығы. Аталған жайт көптеген мүмкіндіктердің ішінен ең дұрысы деп саналатын біреуін ғана тандап алу ықтималдығын күшейтеді.

3) лингвистикалық мәліметтерді сақтайтын қатал алгоритмнен барынша «жұмсақ» жүйеге ауысу мүмкіндігі. Соңғы жүйеде барлық лингвистикалық мәліметтер олардың алгоритмінен бөлектелген (Лейкина және т.б. 1986; Кулагина, 1979).

Жоғарыда аталған жүйелердің сипаттамалары мен ерекшеліктері машиналық аударма жұмысын ұйымдастыруға он әсерін тигізді. Басқаша айтқанда, олар тілшілер мен математика мамандарына жүйенің ерекшелік аспектілерін қарқынды түрде зерттеуге бағыт-бағдар беріп, ынталандыруға себепші болды. Сонымен бірге аталған жайт тілдің формалды моделін құру жолын зерттеу мен оны дұрыс қолдану мүмкіндіктерін анықтауды қажет етті. Ал табиғи тілді формалды сипаттау әдісі грамматикаларды бірізділікке келтіруге мүмкіндік тудырды. Сондықтан машиналық аударманың даму кезеңінің екінші сатысында әртүрлі тілдерді сипаттау үшін формалды модельдер қолданумен бірге тілдік деректерді қалаған дәлдікпен (адекватно) сипаттауға мүмкіндігі мол жаңа модельдер үлгісін құру қажет болды.

Сонымен, машиналық аударманың даму кезеңінің екінші сатысы бірінші буынға қатысты практикалық жүйелердің дамуымен, тәжірибе жинақтаумен, оны іске қосумен, сөздік базасының кең түрде өріс алуымен және тәжірибелік тұрғыда болса да барынша жабдықталған теориялық жүйелерді құрумен сипатталады.

Қол жеткен практикалық және теориялық нәтижелер машиналық аударманың даму сатысының келесі үшінші кезеңіне көшуге даяр екендігінің айғағы болып саналды. Бірақ келесі сатыға өту жолы бірнеше экстралингвистикалық деректерді алдын ала анықтап алуды қажет еткені белгілі.

Біріншіден, 70-ші жылдардың аяғында ақпарат ағымы бұрын-соңды болмаған қарқында дамиды. Мысалы, америкалық және канадалық корпорациялар бір жылдың ішінде 2,5 млрд. сөзге дейінгі мәтіндерді аударып отыруға мүмкіндігі болған. Егер АҚШ-та 1966 ж. аударма жұмысының біржылдық шығыны 22 млн. доллар болса, 1981 ж. ол шығын 200 млн. долларға дейін өседі. Кеңес Одағында штаттағы аудармашылардың айлығын санамағанда, тек тапсырыс бойынша орындалатын аударма жұмысының өзіне орасан үлкен қаржы жұмсалған екен. Бірақ ақпарат ағымының (ғылыми-техникалық мәтіндер, патенттер, құжаттар мен арнаулы қағаздар) соншалық үлкен дәрежеде болуына байланысты оларды тек дәстүрлі жолмен, яғни аудармашылар күшімен аударып шығу мүмкін емес еді. Сол

себеппті, қазіргі кезде автоматтандыру арқылы техникалық стильдегі мәтіндерді аударуды жеңілдету мен жетілдіру маңызды мәселеге айналып отыр.

Екіншіден, спеллер типті өндірістік лингвистикалық процессорларды құрастыру мен оларды баспа мен лексикографиялық жұмыстарда іс жүзінде қолдану, сол сияқты сөздерді өңдеу жүйесінде (word processors) күшті терминологиялық банктерді пайдалану — «лингвистикалық» автоматтанған жүйелерді қолдану арқылы нақты практикалық нәтижелерге ие болу мүмкіндігінің кепілі.

Үшіншіден, машиналық аударманың жаңа кезеңіне өтуге қатысты ерекше жасалған алғы шарт ретінде есептеу техникасының дамуына байланысты барынша жағдай жасалуы — дисплейлердің пайда болуы, дербес компьютерлер мен сканерлердің өмірге келуі, дербес ЭЕМ-ның күш-қуатының өсуі және т.б. Міне, осының бәрі теоретика-лингвистикалық зерттеулерге көзқарасты басқаша өзгертіп, оны практикалық деңгейге дейін көтерді.

Машиналық аударма жүйелерін құрудың бірінші және екінші кезеңдерін сипаттауда, біз нақты лингвистикалық ұжымдардың жұмыстарына айрықша назар аудардық. Ал енді машиналық аударманың қазіргі кезеңін сипаттау үшін оның стратегиясын тандау мен практикалық жүйелерін тиімді пайдалануға байланысты негізгі теоретикалық және техникалық мәселелерді анықтауға қатысты кейбір мәселелерге тоқталайық. Әрине, машиналық аударманың мұндай негізгі мәселелерін талқылау машиналық аударма жүйесінің болашақтағы даму жағдаятын да сөз етуді қажет етеді.

Бірінші мәселе мынадай мүмкін болатын үш түрлі теориялық стратегияларды талқыға салуды қажет етеді:

- тікелей бинарлық машиналық аударма;
- трансфер арқылы аудару;
- аралық-тіл арқылы аудару, сонымен бірге олардың грамматикалық мүмкіндіктерінің ерекшеліктерін іске асыру.

Тікелей бинарлық аударма машиналық аударманың «ескі» жүйелерінен орын алса да, сол қалыпта машиналық аударманың көптеген практикалық жүйелерінің негізгі стратегиясына айналып отыр. Мұндай тәсіл нақты тілдік жұпқа, яғни

қос тілдерге бағышталады (GAT, SYSTRAN жүйелері және олардың дериваттары Бүкілодақтық аударма орталығында зерттелген АРМАС жүйесі). Бинарлық аударма әрекеті кезінде сегменттердің шекарасын анықтау мен омонимияны мәнмәтінге қатысты ажырату негізінде сөйлемдерге қарапайым өзгертулер (трансформация) жүргізіледі.

Бұрынырақ зерттелген машиналық аударма жүйесін құрастыру тәжірибесі мен оны пайдалану тиімділігі, олардағы аса күшті сөздіктер қоры мен арнайы түзету (редакциялау) тәсілдерінің негізінде деуге болады. Оған Джорджтаун университетінің GAT жүйесі мен оның түрлі варианттық жүйелері мысал бола алады. Бірақ жеке лексика-морфологиялық және семантика-синтаксистік ішкі құрылымдар (подструктуры) негізінде іске асатын тікелей бинарлық сәйкестендіру жүйенің лингвистикалық және программалық жабдықталуын тек нақты тілдер жұбының ерекшеліктеріне ғана бағыштап қоймай, сонымен бірге ішкі тілдердің (подъязык) айырым-белгілеріне де қатаң түрде көңіл бөлуді қажет етеді. Егер тіл жұптарының бір сыңары басқа бір тілге ауыстырылса, іле-шала лингвистикалық, ал кейбір жағдайларда программалық жабдықтауларды да қайта құру қажет болады.

Трансформациялық машиналық аударма (трансфер арқылы аударма) әдісі де машиналық аударма дамуының екінші кезеңінде қолданылып, сынақтан өтті. Трансферді пайдаланатын бұл стратегияны қолдану – түрлі тілдердің мәтіндеріне тікелей қатысы бар жазбаларды жүзеге асырудың мүмкін еместігі жөніндегі теориялық болжамға негізделеді. Мұндай байланыс түрі әртүрлі деңгейдегі жалғас жұмсалымдар жолымен анықталуы мүмкін: лексикалық, морфологиялық, синтаксистік құрылымдар деңгейінде (конфигурациялық және реляциондық синтаксис терминдерімен) және синтаксис деңгейінде сипаттау. Түрлі деңгейдегі талдаулар кезінде «кіретін» сөйлемнің құрылысын сипаттайтын трансферді пайдалану бағыныңқы таралымнан базалық грамматикалық бірліктер ретіндегі басты астырт құрылымға (к глубинной ролевой структуре) біртіндеп өтуді көздейді және одан әрі «шығатын» бағыныңқы таралымға өтеді. Мұндай кезде тек белгілі бағыныңқы таралымға қатысты қайта құру әрекеті сөйлемнің толық түсінігінің моделін құрастырмай-

ақ машиналық аударма жасауға мүмкіндік тудырады. Бұл мәселенің шешімін табу машиналық аударма кезінде міндетті емес, өйткені астырт (глубинной) семантикалық-синтаксистік құрылымдарға түсініктеме беру тұтынушының үлесіне қалдырылады, яғни ЭЕМ-нан шыққан аудармамен әрі қарайғы жұмыс істейтін маман аудармашының үлесіне тиеді. Трансферді қолдандың мұндай тәсілі көптеген шетел ұжымдарымен бірге машиналық аударма жүйесін қолданатын бірқатар ресейлік ұжымдарға да тән болып келді.

Аталған жүйелерде аударманың негізгі үш сатысын бөліп айтуға болады: талдау (анализ), трансфер және жинақтау (синтаз). Әрбір саты жүйелі субпроцестер түрінде орындалады. Мәселен, жоғарыда аталған SUSY атты жүйеде талдау әрекеті 8 субпроцесті қамтиды. Нәтижесінде талдауға қатысты «кіретін» сөйлемнің таралымы құрастырылады және «шығатын» сөйлемнің таралымы жайлы мәлімет 8 субпроцессордан тұратын трансфер түрінде көрініс табады. Талдау кезеңінде бұл таралым үш субпроцестің өзара әрекеттесуінің нәтижесінде аралық-тілдегі сөйлем түріне қайта келтіріледі. Осыған ұқсас тәсілдер трансфермен әрекет ететін басқа да жүйелер түрлерінде қолданылады: CADA, GETA, TAUM-METEO, жапон жүйесі, EUROTRA жүйесі, Ресейдегі СИЛОД (MULTIS) жүйесі, А.И.Герцен атындағы РГПУ-да зерттелген ФРАП, АРАП, ЭТАП, ЭТАП-2, STYLUS және т.б.

Еуропалық экономикалық бірлестіктің айрықша комиссиясы жанында 1982 жылы машиналық аударманың дамуының үшінші кезеңінде ұсынылған EUROTRA жүйесінің негізі қаланды. Аталған жүйенің тілдері: ағылшынша, французша, немісше, итальянша, испанша, португалша, голландша, датша, жаңагрекше. Бұл жоба БЭЖ (ЕЭС) елдерінің тоғыз ресми тілі үшін машиналық аударма жүйесін құруды және 72 тілдің жұбы бойынша аударма жасау қарастырылды.

Аталған тілдердің қайсысын болса да аударуға қажетті талдауда интерфейстік көрініс деп аталатын синтаксистік қатынастардың семантикалық түсінігі беріледі. Мәтін семантикалық бағыныңқылықтардың түзілістері арқылы көрініс табады. Мысалы, олар астырт септігі типінде (типа глубинных падежей) болуы мүмкін. Бұл жерде арнайы атап кетуге

тұрарлықтай жайт – бір тілдің интерфейстік құрылымының басқа тілдегі осындай құрылыммен сәйкес келмеуі және осыған қарамай трансферді қолдануға қажетті мүмкіндіктің туатындығы. Сонымен, ана тілінің аясында аударма жасайтын әрбір маман талдау мен жинақтау тәсілдерінен басқа трансфердің 8 түрлі процедурасын да құра білуі қажет.

Жобаның мақсаты – өндірістік машиналық аударма жүйесінің бастапқы үлгісін жасау. EUROTRA жүйесінде іске асатын трансфердің ерекшелігі, оның қадамдап әрекет етуі. Әрбір қадамды жасаудан бұрын алдыңғы қадамдағы әрекеттердің бұрыс нәтижелері «тазартылады».

Трансферді пайдаланатын жүйелерге ЭТАП-2 атты жүйе де жатады (*Апресян и др. 1989*). Бұл жүйе ағылшынша-орысша машиналық аударма жасай алады. Бұл идеологиялық тұрғыдан қарастырғанда ИНФ-ЭЛ жүйесінің жетілген түрі және айрықша сөздіктермен жабдықталған – нақты бір пәндік салаға арналған, ұластыруға негізделген (комбинаторлық) автоматты сөздік. Зерттеушілер мұндай «тәуелсіздік» мүмкіндігін тілдік модельдің жан-жақты зерттелуімен байланыстырады.

Трансфердің барлық ережелері жүйе ішінде жалпылық, жекелік (трафареттік) және сөздіктік деп үшке бөлініп қарастырылады. Жүйеде 45 жай (элементарлық) предикаттар мен 5 синтаксистік қатынастар орын алған. Сүзбеден өткізу тетігі (механизмі) мен жоғары баға беру ережелері ұсынылған синтаксистік болжамдар ішінен тек біреуіне ғана артықшылық беруге мүмкіндік туғызады.

Трансферді қолдану талдаудың әртүрлі деңгейінде «кіретін» сөйлемнің құрылымын сипаттайтын ішкі тармақтардан (поддережья) базалық грамматика берілісін қолданатын астырт басты құрылымға (к глубинной ролевой структуре) жүйелі түрде көшу жолы мен одан әрі қарай «шығатын» ішкі таралымдарға көшу әрекеті қарастырылған. Осындай жағдайда белгілі ішкі тармақтар үшін қайта құру әрекеттері сөйлем түсінігін модельдемей-ақ, машиналық аударма жасауға мүмкіндік тудырады. Ұғынуды модельдеу міндеті машиналық аударма мәселесінен тыс жатады, өйткені машиналық аударма мәселесіндегі астырт (глубинный) түсініктеме алу адам үлесіне, яғни нақты пәндік сала маманына жүктелетіні жоғарыда да аталды.

Трансфер арқылы машиналық аударма жүйелерін іске қосатын мұндай тәсіл машиналық аударма жүйесі дамуының екінші кезеңінде пайда болған СИЛОД жүйесінде де қолданыс тапты. Қазіргі кезде SILOD-MULTIS көптілді жүйе дербес компьютерлерде аударма жұмысын атқарады. Машиналық аударманың барлық процесі лексика-морфологиялық және семантика-синтаксистік субпроцестердің композициясы ретінде SILOD-MULTIS жүйесінде модельденеді. Лексика-морфологиялық талдау кезінде мәтіндер мен сөздік бірліктерін сәйкестендіру жұмысы жүргізіледі және сонымен бірге автоматты сөздік топтамасынан сөздікке қатысты ақпараттар шығарылып алынады. Мұндай талдаудың нәтижесі ретінде мәтіннің кесте түріндегі көрінісін және канондық пішіндегі аударма баламалары (переводные эквиваленты) жайлы ақпаратты айтуға болады. Бұл кестеде әрбір сөзтұлғаға оның лексика-семантикалық сипаттамасы сәйкестікке қойылады. Сәйкестіктің дәлдігі конверсиялық және/немесе омонимдік септелуге (подежная омонимия) дейін, функционалдық, басты (ролевой) және семантикалық сипатына дейін анықталады. Мәтінді талдаудың барлық деңгейлеріндегі берілісі сияқты, аталған көріністер әртүрлі жағдайда қолданылуы мүмкін (машиналық аудармада, аннотациялауда, рефераттауда және т.б.). Егер машиналық аударма жайлы ғана айтатын болсақ, онда талдаудың бұл сатысында алынатын нәтиже — лексика-морфологиялық және фразеологиялық тұрғыда жолма-жол (подстрочник) аударылған мәтін.

Семантика-синтаксистік өңдеудің нәтижесінде құрылымды тіларалық деңгейде қайта құру жүзеге асады. Ол үшін лексика-морфологиялық талдау сатысындағы барлық ақпарат іске қосылады, одан кейін барып грамматикалық және семантикалық топтамалар пайдаланылады. Мұндай трансферде иерархиялық бөліктер (компоненттер) мәтін ішінен алынатын тік қисындасқан жүйешіктердің жиынтығы ретінде модельденеді.

Семантика-синтаксистік талдау сөйлем құрылымының әртүрлі деңгейлерінде жүйелі түрде жүзеге асады: топтар деңгейі, функционалды сегменттердің деңгейі және, ең соңында, сөйлем деңгейінде (Беляева, 1986; Апресян және т.б., 1989, Соколова, 1990). Топтар деңгейінде топтардың шекаралары мен лексикалық бірліктің өзара байланысы белгіленеді. Келесі

деңгейде бөлініп алынған тізбектер сөйлемнің тұрлаулы мүшелеріне сәйкес келетін функционалды сегменттерге біріктіріледі. Бөлініп алынған сегменттерде мәтінді кестелік көрініске модификациялау деңгейінде қайта құру әрекеті жүзеге асады. Функционалды сегменттер деңгейінде алынған нәтиже — шектеулі грамматика бойынша жасалған машиналық аударма болып саналады. Толық түрдегі семантика-синтаксистік талдау сөйлем деңгейінде ғана іске асады.

Талдау мен трансфер процедураларын жүзеге асырғанда синтаксистік сипаттау маңызды рөл атқарады, өйткені тек осының негізінде ғана сөйлемнің мүмкін болатын фреймі танылады. Мұндай таным әрекетінде актанттар жиыны, менгеру моделінің мүмкіндігі жайлы және т.б. ақпараттарды бойында сақтайтын етістіктің лексика-синтаксистік сипатталуы негізгі рөл атқарады.

Осы әдістемеге сәйкес SILOD-MULTIS машиналық аударма жүйесі автоматты сөздіктер мен грамматикалардың жинағы ретінде жүзеге асады. Жүйенің негізгі пішін үйлесімділігінде үндіеуропа және басқа құрылымдық тілдермен әрекет ету мүмкіндіктері алдын ала қарастырылған. Мұндай мүмкіндік барлық «кіретін» тілдердің сөзтізбелерінің әмбебаптық құрылымына және «шығатын» орыс тілінің бірыңғай құрылымына негізделеді.

Нақты әрекеттегі машиналық аударма жүйелерінің трансферді пайдалануы дәл осы стратегияны таңдаудың мақсатқа сай екендігін дәлелдей түсті. Сонымен бірге жасанды интеллект жүйесінің даму деңгейі мен есептеу техникасының жаңа мүмкіндіктері машиналық аударманың үшінші стратегиясына қызығушылықты жаңаша түрде жандандыра түсті.

Аралық-тілді қолдануға қатысты аударма стратегиясы машиналық аударма прагматикасына білім базасын құрудағы машиналық аударма мен оның идеологиясының әсерін бейнелейді. Машиналық аударманың даму сатысының бірінші кезеңінде қолданылатын аралық-тіл тәжірибе жүзінде ғана жүзеге асқанымен, ол өз уақытынан айтарлықтай озды — оның пайда болуы машиналық аударманың жаңа кезеңі деуге болады. Мысалы, аралық-тілді қолдана отырып, Карнеги университетінде CMU жүйесі, Утрехте зерттелген DLT жүйесі, Японияда —

JETR, АҚШ-та – SAM және TRANSLATOR жүйелері машиналық аударма жасауда қолданыс тапты.

Аралық-тілді пайдаланатын машиналық аударма жүйесі мен жасанды интеллект жүйесін бір-біріне қисындастыру мәселесі – сөйлем мағынасын немесе тұтас мәтіннің мазмұнын тануға және оның мазмұнын автоматты түрде сөйлеу тілінде туындату үшін күрделі білім базасын қолданудың қажеттігімен түсіндіріледі.

Жасанды интеллект жүйесінің аралық деңгейі болып саналатын ұғынудың (понимание) концептуалды деңгейін модельдеу бірнеше базалық құрастырушыларға негізделеді:

- ұғым түйіндері (узлы-понятия) мен байланыс типтері жайлы сөз болатын мақала түріндегі сөздіктерді құрастыруды қажет ететін білімді тануға көмектесетін тілдер;

- «белгі – мағына» принципі бойынша қайта құруды жүзеге асыратын жүйелі грамматикалар.

Қазіргі кезеңдегі аралық-тілдің негізі ретінде білімді танытатын тілдер саналады. Осыған қатысты талдаудың мақсаты «кіретін» сөйлем бойынша алынатын және білім базасының ақпаратымен толықтырылатын сөйлем мағынасын таныту. Аударманы жүзеге асыру әрекетінде «кіретін» тілдің құрылымын «шығатын» тілдің құрылымына сәйкестендіріп, қайта құру қажет болады. Әрине мұндай аударма нәтижесі шын мағынасындағы аудармадан гөрі, берілген мәтіннің мазмұнын не сөйлемнің мағынасын қайталап айтып беру жағдайына көбірек ұқсайды.

Осындай жағдаятта білімді танытатын тілді аралық-тілдің өзі деуге болады. Мұндағы аралық-тіл «кіретін» фразаның (сөйлемнің) семантикалық желісі мен «шығатын» фразаның ішкі берілісінен қайта туындауға қатысты семантикалық таңбалауды («разметка») жүзеге асырады. Осылайша қайта туындату (генерация) белгілі үлгі бойынша (шаблон, фрейм және т.б.) орындалады. Бұл әдістің жалпытеориялық тартымдылығына қарамай, ол тек тәжірибелік деңгейде ғана жүзеге асты. Себебі оның автоматтандырылған лексикондарының негізінде аса күшті семантикалық желі құруға қажетті лингвистикалық негіздемесі мен программалық құралдары әлі де болса толық қамтылмаған еді.

Машиналық аударма жүйелерінің кең түрде қолданыс табуы оның эксперименттік дәрежесінен автоматтанған дәрежесіне, редакциялау ерекшелігіне, мақсатқа сай келуіне қарай аудармашының жұмыс орнына айналуына жағдай жасайды (*Machine Translation, 1987; Апресян және т.б., 1989*).

Машиналық аударма жүйелерінің қазіргі сатыдағы даму ерекшеліктерін қарастыра келіп, оны әр кезеңге бөліп топтап, баға беру дұрыс бола бермейді. Себебі машиналық аудармаға қатысты ондай топтаулар есептеу машиналарының (компьютерлердің) кезеңді өзгерістеріне және мүмкіндіктерінің артуына тікелей байланысты.

Қазіргі кездегі машиналық аударма жүйелерінің зерттелуі мен қолданылуы мына жайтты байқатады. Машиналық аударма жүйелерін талдау және олардың функционалдық критерийлері бойынша салыстыру мынаған саяды:

1. Жүйенің прагматикасы. Машиналық аударма жүйесін өндірістік немесе эксперименттік дәрежеде құруға байланысты оның негізгі сипаттамасы мен талдау әдістемесін тандау түрлері ерекшеленеді. Бұл өлшем сөздік базасының көлеміне қойылатын талаптарды (өндірістік машиналық аударма жүйесінде бір тілдер жұбына 20000 сөздік мақаладан кем болмау) анықтайды. Мысалы, Bravice – Weidner LOGOS, GAT, S'YSTRAN және т.б. жүйелерде сөздік мақала көлемі көрсетілген мөлшерден кем емес. Бұл шартқа бағынбайтын жалғыз жүйе – TAUM-METEO. Оның сөздік мақала көлемі 1500-ге тең, өйткені бұл жүйедегі «кіретін» мәтіндердің тақырыптары қатаң түрде шектелген. Тәжірибелік (эксперименттік) жүйелерде автоматты сөздіктер «төменгі» жағынан шектелмеген, яғни көлемі 200 сөздік мақаладан тұратын сөздіктер бойынша да ол жүйе аударма жұмысын атқара алады.

Өндірістік машиналық аударма жүйесін құрудың мақсаты анық – үзіліссіз процесс түрінен ұйымдастырылған құжат ағымдарынан машиналық аударма алу.

Машиналық аударманың өндірістік жүйелері мынадай талаптарға сай келуі керек:

– **Жүйе жұмысының тұрақтылығы.** Машиналық аударма жүйесінен алынатын нәтиже сапасының ойдағыдай болмауына қарамай, тұтынушы оны өз қажеттігіне пайдалана алатындай

мүмкіндігі болуы керек. Машиналық аударма жүйесінен «шығатын» ақпараттың сапасының төмен болуы, көбінде, жүйенің сөздіктермен толық жабдықталмауына байланысты болып келеді. Сондықтан жүйеден «шыққан» кейбір сапасы төмен («дөрекі») аудармалар әрі қарай маман аудармашының қолынан қайта түзетіліп, сапасы жақсартылады. Осыған байланысты SYSTRAN жүйесін пайдалану тәсілімен танысу қажет.

– **Жүйенің тираждануы (таратылуы).** Аударма жүйесінің қолдану аясы кең болуы үшін оның программалық және лингвистикалық құралдары барынша қарапайым түрде болуы қажет. Сонымен бірге жүйеде басқа тақырыптық аяға және тілдерге көшу әрекетін жеңілдетуге қажетті әмбебаптық модельдер орын алуы керек.

– **Жүйенің икемделінуі (адаптациялануы).** Машиналық аударма жүйесінің нақты тұтынушының мұқтаждығына және өңделетін материалдың ерекшелігіне қарай икемделу мүмкіндігі болуы қажет.

– **Уақытша параметрлердің оптималдығы (аукымдылығы).** Машиналық аударма жүйесінің аударма жасау жылдамдығы уақыт бірлігі ішінде қабылданатын ақпарат көлемімен немесе жүйені тұтынушы ретіндегі редакторлардың, аудармашылардың, мамандардың жұмыс әрекетімен сәйкес келіп отыруы қажет.

– **Тұтынушыға жайлылық (комфорттық).** Жүйенің сервистік құралдары тұтынушының әрекеттерінің «пакеттік» және диалогтік режимдері үшін аса қолайлы болуы шарт.

Эксперименттік жүйелерді құрудың мақсаты машиналық аударманың даму жолдарын зерттеу, адамның ойлау-сөйлеу қызметін модельдеу болып табылады. Тек осы аталған тұрғыда ғана машиналық аударма мен жасанды интеллектің міндеттері бір-бірімен сәйкес келе алады. Эксперименттік жүйелер мен оларды құрастырушы ғалымдар машиналық аударманың көшбастар ролін атқарады да, ал олардың қол жеткізген нәтижелері өндірістік жүйелерде қолданыс табуы мүмкін. Мысалы, Саарбрюкен университетінің SUSY атты эксперименттік жүйесін құру арқылы алынған нәтижелері бойынша MARIS

атты өндірістік жүйе құрылуда және SUSANNAH атты аудармашының автоматтанған жұмыс орны зерттеліп, жасалуда.

2. Жүйенің автоматтану дәрежесі. Практикалық тұрғыдан қарастырғанда, машиналық аударма жүйелерінің барлығы да адамның (маманның) араласуын қажет етеді. Бірақ бұл өлшемді қолдану жүйені екі қарама-қарсы көзқараста топтауға мүмкіндік береді: редакциялауда қажет ететін машиналық аудармадан бастап, аудармашы-адамға көмек беретін жүйеге дейін (*аудармашының автоматтандырылған жұмыс орны жайлы Бүкілодақтық аударма жасау орталығы құрастырған СПАС атты жүйені қара*).

Шынында да, бұл жүйелердің негізгі айырмашылығы адамды аударма жұмысына тарту стратегиясы – жүйе арқылы алынған нәтижені редакциялау немесе машинасыз (дәстүрлі) аудару кезінде машиналық аударма жүйесінің аса күшті сөздіктер мүмкіндігін пайдалану. Осыған қарай редакциялау кезеңі мен оның көлемі айырым табады. Мысалы, машиналық аударманың өндірістік жүйелерінде аударылған материалдың 20 пайызын қамтитын бөлігі үшін соңынан редакциялау (постредактирование) тәсілі кең түрде қолданылады (*ALPS, ENGSPAN, LOGOS, METAL, SPANAM жүйелерін салыстыр*). Әсіресе, бұл жайт мәтіндер жиынын пакеттік режимде өндеуді қарастыратын жүйелерге тән болып келеді.

Машиналық аударма нәтижесі, ең алдымен, әрі қарайғы өңдеулерді қажет ететін бастапқы аударма нәтиже болатындығын ескере отырып, машиналық аударма мәселесін зерттеуші ғалымдар арнайы сервистік жүйені құрудың және практикалық машиналық аударма жүйесінің нәтижелігін бағалау үшін ондай сервистердің деңгейлік дәрежесінде де ескеру қажет екендігін байқауда. Мұндай сервис, әдетте, редакциялау құралдарын, сөздік жүйелерін интерактивті түрде енгізу құралдарын (автоматтанған лексикондарды) және т.б. өз құрамында сақтайды.

Ақпарат ағымының жаңа толқыны машиналық аударма жүйелерін күнделікті өмірде кең түрде қолдану қажеттігін айқындап отыр, сондықтан қажетті ізденістер барысында жоғарыда аталған мәселелер толық түрде ескерілуі керек.

Э д б и е т

Автоматический перевод. 1949–1963 гг.: Критико-библиографический справочник / Сост. И.А. Мельчук, Р.Д. Равич. М., 1967. 516 с.

Автоматический перевод: Сб. статей / Под ред. и предисл. О.С.Кулагиной, И.А.Мельчука. М., 1971. 357 с.

Андреев Н.Д. Машинный перевод и проблема языка-посредника // Вопр. языкозн. 1957 №5. С. 117-121.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы «ЭТАП - 2». М., 1989. 295 с.

Бар-Хиллел И. Будущее машинного перевода // Науч. докл. высшей школы. Филол. науки. 1962. №4. С. 203-206.

Белопольская А.Р., Ковригин А.Б. Исходный этап независимого анализа немецкого простого распространенного предложения // Материалы по математической лингвистике и машинному переводу. Вып. 2. Л., 1963. С. 150-160.

Беляева Л.Н. Применение ЭВМ в лингвистических исследованиях и лингводидактике. Л., 1986. 84 с.

Беляева Л.Н., Откупщикова М.И. Автоматический (машинный) перевод // Прикладное языкознание: Учебник. Отв. ред. А.С. Герд. СПб.: Изд-во С. – Петербург. ун-та, 1996. 528 с.

Воронин В.А. Система независимого конфигурационного анализа китайского текста при машинном переводе // Машинный перевод. М., 1961. С. 135-172.

Дрейзин Ф.А., Рашитов Р.С. Принцип синтаксического анализа татарской фразы // Машинный перевод. М., 1961. С. 295-303.

Кремнева Н.Д., Морева В.Ю., Ульянова Г.С. Принципы построения грамматики русского синтаксического анализа с учетом предпочтения и порядка замыкания связей // Лингвистические проблемы функционального моделирования речевой деятельности. Вып. 3. Л., 1976. С. 139-144.

Кулагина О.С. Исследования по машинному переводу. М., 1979. 319 с.

Кулагина О.С., Мельчук И.А. Машинный перевод с французского языка на русский // Вопр. языкозн. 1956. №5. С. 111-121.

Лейкина Б.М., Никитина Т.Н., Откупщикова М.И. и др. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ // Науч.-тех. информация. 1996. № 1. С. 40-50.

Леонтьева Н.Н. Система французско-русского перевода (ФРАП): лингвистические решения, состав, реализация // Машинный перевод и

прикладная лингвистика: Проблемы создания системы автоматического перевода. Вып. 271. М., 1987. С. 6-26.

Марчук Ю.Н. Проблемы машинного перевода. М., 1983; Методы моделирование перевода. М., 1985. 199 с.

Машинный перевод: Сб. статей. М., 1957. 314 с.

Мельчук И. А. Работы по машинному переводу в СССР // Вестн. АН СССР, 1959. № 2. С. 43-47; Опыт теории лингвистических моделей «смысл – текст». М., 1974. 314 с.

Молошина Т. Н. Алгоритм машинного перевода с английского языка на русский // Проблемы кибернетики. Вып 3. М., 1960. С. 209-272.

Откупщикова М. И. Синтаксис связного текста. Л., 1982. 103 с.

Панов Д. Ю. Автоматический перевод. 2-е изд. М., 1958. 72 с.

Переводная машина в Таллине в 1924 г. // Сообщения по машинному переводу. Сб. 1. Таллин, 1962. С. 95-96.

Переводная машина П. П. Траянского: Сб. матер. о переводной машине для перевода с одного языка на другие, предложенной П.П.Троянским в 1933 г. / Отв. ред. Д. Ю. Панов. М., 1959. 52 с.

Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л., 1979. 112 с.

Синтаксический компонент в системах машинного перевода // ВЦП. Сер. 2. Машинный перевод автоматизация информационных процессов: Обзорная информация. Вып. 5. М., 1981. 130 с.

Шаляпина З. М. Англо-русский многоаспектный автоматический словарь (АРМАС) // Машинный перевод и прикладная лингвистика. Вып. 17. М., 1974. С. 7-67.

Шумилина А. Л. Вопросы анализа личных местоимений третьего лица // Лингвистические исследования по машинному переводу. Вып. 2. М., 1961. С. 132-140.

Цейтин Г С., Откупщикова М. И., Лейкина Б. М. и др. Система анализа текста с процедурным представлением словарной информации // Актуальные вопросы практической реализации систем автоматического перевода. Ч. 1. М., 1982. С. 205-231.

Чижиковский В.А., Беляева Л.Н. Тезаурус в системах автоматической переработки текста. Кишинев, 1983. 163 с.

Hutchins W. J. Transiation: Past, Present, Future. Harwood, 1986.

Machine Transiation: Theoretical and Methodoiogikai Issues // Ed. by S. Nirenburg. New York, 1987. 350 p.

Баспа ісін автоматтандырудың лингвистикалық негіздері

Баспа ісін автоматтандыру тек авторлық шығарманың түпнұсқасын полиграфиялық түрге келтіруді электронды есептеу машинаға (ЭЕМ-ға) толығымен жүктеп қана қоймайды, сонымен бірге бірқатар лингвистикалық өңдеу әрекеттерін автоматтандыру мәселесімен де айналысады. Баспа ісіндегі мәтінді өңдеуге қатысты лингвистикалық жабдықтау кезеңдерінің сипатын қысқаша түрде баяндайық.

Баспа ісін автоматтандырудың кезеңдері. Қазіргі кезде мұндай әрекеттің мынадай кезеңдерін бөліп қарастыруға болады:

1) **авторлық кезең** (мәтін авторының шығармасын ақпарат сақтайтын тетікке көшіруі);

2) **редакциялық кезең:**

а) авторлық түпнұсқа мәтінге корректура (түзету) жүргізу мен редакциялау;

ә) редакциялық баспа қызметі (рецензиялау, корреспонденцияларды өңдеу, деректер базасынан ақпарат іздеу және т.б.);

3) **құрылымдау кезеңі** (мәтіннің әр элементіне полиграфиялық параметрлерді белгілеу: терме мен бет пішімі, шрифтердің кегелі мен гарнитурасы және т.б.);

4) **полиграфиялық қайта туындату (воспроизводство) кезеңі** (мәтіннің әр элементі үшін белгіленген полиграфиялық параметрлерді фототерім машиналары арқылы сыннан өткізу);

5) таралымды белгілеу кезеңі, яғни мәтінді ұсынылған дана санына сәйкес басып шығару (егер басылым дәстүрлі қағаз түрінде болса);

6) тарату кезеңі (басылымды ақпарат сақтайтын тетік арқылы немесе байланыстың электрондық торабы арқылы баспа қорынан оқырманға ұсыну).

5-кезенде тілдік деректер толығымен қатыспайды, ал басқа кезендерде оның тек орташа дәрежесінде ғана қатысы бар. Тілдік деректердің ең күшті қажеттілігі – 2-кезең.

Баспа ісін автоматтандыру кезендері хронологиялық жағынан бір-біріне тұтасып, бірігіп кетуі мүмкін. Мысалы, шығарма авторы мәтінді өзі теріп, алғашқы түзетулерді өзі жүзеге асыруы да мүмкін. Сондықтан лингвистикалық жабдықтау мәселесі баспа ісін автоматтандырудың әр кезеңінде де кездеседі.

Енді жоғарыда аталған кезендердің әрқайсысына байланысты лингвистикалық жабдықтаудың ерекшеліктеріне тоқталайық.

Авторлық кезең. Бұл кезеңде мәтінді теру үшін авторлар әртүрлі мәтіндік процессорларды қолданады, яғни мәтінді теру мен оған өзгертулер жүргізуде компьютерлік программаларды пайдаланады. Мұндай мәтіндік процессорларда тілдің мынадай қызметтері орын алады:

1) автордың қалаған сөзіне синонимдерді іздестіру және қажеттікке қарай, ол сөзді автоматты түрде оның синонимімен алмастыру;

2) қате жазылған сөздерді эталондық орфографиялық сөздіктен іздестіру немесе ол сөздің сөздікте жоқ екендігін айқындау;

3) тұлғалық сәйкестік бойынша эталондық сөздіктегі сөзді мәтіндегі тиісті орнына көшіру.

Аталған тілдік қызметтердің біріншісі осы авторлық кезеңде жүзеге асады, ал екіншісі мен үшіншісі келесі кезендердің негізін құрайды, сондықтан оларды кейінірек қарастырамыз.

Авторлық кезеңде қалам иесі қажетті деген мәтін элементтерін бөліктерге бөліп алуы үшін таңбалар жүйесінің арнайы жазылған формалданған тілін қолдануына мүмкіндігі бар. Егер автор ондай таңбаларды шығарма мәтініне енгізбей

кетсе, баспа жұмысын құрылымдау кезеңінде ол міндетті техникалық редакторлардың атқарулары қажет.

Сонымен бірге бұл кезеңде автордың машиналық ақпарат сақтайтын тетікте жазылған түсіндірме және энциклопедиялық сөздіктерді автономдық сипатта қолданылуына да мүмкіндігі бар. Жүйенің программалық жабдыкталуы тұтынушының өзіне қажетті деген сөздік мақаланы тез тауып алуына жағдай жасайды.

Редакциялық кезең. Бұл кезеңде баспа ісін автоматтандыру жүйесі мәтінді өңдеуге қатысты корректура жүргізу мен редакциялау қызметтерін жүзеге асырады. Мәтінді өңдеу жұмысымен қатар редакциялық-баспа қызметі де орындалады. Аталған қызметтер осындай мақсаттарға арнайы арналған редакциялық-баспа жүйесінде және баспа ісін автоматтандыруда қолданылатын басқа да жүйелерде жүзеге асады.

Баспа ісін автоматтандырудағы корректуралау мен редакциялау нысаны ретіндегі шығармалар мәтіні мынадай элементтерден тұрады:

- жәй ғана мәтіндер (графемалар, яғни әріптер, морфемалар, сөздер, сөз тіркестері, сөйлемдер, фразадан тыс тұтастық; топтамалар, яғни бөлімдер, тараулар және т.т.; және дискурс, яғни бүтін түріндегі мәтін);

- күрделі мәтіндер, яғни математикалық өрнектер мен кестелер;

- шығарма аппаратын құрастырушы мәтіндер (титулдық бет, басылым мәліметтері, сілтемелер және т.б.);

- қызметтік мәтіндер, яғни мәтін элементтерінің таңбалары және полиграфиялық безендіруге қатысты бұйрықтар.

Редакциялау кезеңінде баспа ісін автоматтандыру әрекетіне қатысты авторлық мәтінді өңдеудің мынадай екі түрлі амалын атауға болады:

- 1) мәтіндерді түпнұсқаға сәйкестендіру;

- 2) мәтіндерді қазіргі кездегі тілдік нормаға сәйкестендіру.

Баспа ісін автоматтандыру ісінде бірінші процесті *корректура* жүргізу деп, ал екіншіні — *редакциялау* деп атайды. Корректура жасау мен редакциялау істерінде де екі түрлі әрекеттерді ажыратуға болады:

- а) бақылау амалы, яғни ауытқуларды табу;

ә) қайтақұру (реконструкция) амалы, яғни бақылау барысында анықталған ауытқуларды түзету.

Бұл айтылғаннан мынадай қорытынды шығаруға болады: бақылау амалының нысаны ретінде барлық мәтін көлемі алынады да, ал қайтақұру нысаны ретінде – ауытқулар түрлері орын алатын мәтін бөліктері алынады.

Енді ауытқулардың түрлеріне жататындар:

а) бұрмалаулар (түпнұсқадан айырым табатын ауытқулар);

ә) қателер (нормадан ауытқу);

б) қате жіберу мүмкіндіктері (түпнұсқадан да және нормадан да бірдей кететін ауытқулар).

Әрине, редакциялау барысында бақылауды қажет ететін (баспаның, логикалық, техникалық және т.б.) норма мәселесінің барлығы бірдей емес, тек лингвистикалық және психолингвистикалық түрлері ғана орын алады. Бұл топтардың ішінен мынадай нормалық топ түрлерін бөліп қарастыруға болады:

а) тізімдер (мысалы, неше түрлі сөздіктер, қысқартулар тізімі немесе өлшем бірліктерінің тізімі);

ә) үлгілер (мысалы, сөйлемдер мен қайталаным кезіндегі тыныс белгілерді қою үлгісі);

б) құрылымдар немесе модельдер (мысалы, басылымның құрылымы немесе сөйлем моделі);

в) параметрлер (мысалы, бір бетте қатарынан төрт тасымалдың бірдей кездеспеуі немесе аннотация көлемінің 600 таңбадан аспауы).

Корректурадан немесе редакциялаудан өткен автор мәтіннің сапасын анықтау үшін мынадай шарттардың орындалғаны жөн:

– корректура үшін – мәтін көшірмесінің бұрмалану дәрежесі мына өрнекпен анықталады:

$$E=e/M,$$

мұндағы e – мәтінде кездесетін бұрмалану саны, M – авторлық баспатабақ бойынша есептелген мәтін көлемі;

– редакциялау үшін – қателесу дәрежесі:

$$S=r/M,$$

мұндағы r – мәтіндегі қате саны.

Машиналық (компьютерлік) корректура жүргізу мен редакциялаудың кейбір технологиялық ерекшеліктеріне тоқта-

лайық. Мұндай ерекшеліктердің ең алғашқысы ретінде мәтінді өңдеудің үш кезеңін атап айтуға болады.

Бірінші технологиялық ерекшелікке жататындар:

- корректураның алдын алатын әрекет (предкорректурa) және редакциялаудың алдын алатын әрекет (предредактирование). Мысалы, мәтін элементтерінің әрбіреуіне сәйкестікке қойылатын таңба;

- интеркорректурa (интерредакциялау), яғни компьютер көмегімен орындалатын бақылау мен қайтақұру әрекеттері;

- корректурадан кейінгі әрекет (посткорректурa) және редакциялаудан кейінгі әрекет (постредактирование), яғни техникалық құралдар арқылы орындауға келмейтін жағдайдағы корректор мен редактордың амал-әрекеттері.

Екінші технологиялық ерекшелікке корректурa мен редакциялаудың әр дәрежедегі автоматтану сипаты жатады:

- компьютерленген корректурa және компьютерленген редакциялау (мәтіндік процессор арқылы компьютер жадына жазылған мәтінге адамның араласуы арқылы ғана жүзеге асатын бақылау мен қайтақұру әрекеттері);

- автоматтанған корректурa мен автоматтанған редакциялау (мәтінді бақылау жұмысы ЭЕМ арқылы жүзеге асады да, ал қайтақұру әрекеті – корректор немесе редактор арқылы);

- автоматтанған корректор мен автоматтанған редакциялау (мәтінді бақылау мен қайтақұру жұмыстары ЭЕМ арқылы орындалады, ал адам қалыпты жағдаят бұзылса ғана араласады).

Мәтінді бақылау немесе қайтақұру және де корректурa немесе редакциялау әдістерінің тиімділігін анықтау үшін мына өрнекті қолдану қажет:

$$E=(O_1-O_2)/O_1,$$

мұндағы E – мәтінді өңдеудің тиімділігі, O_1 – мәтінді өңдеуден бұрынғы ауытқулардың саны, O_2 – өңдеуден өткеннен кейінгі мәтіндегі ауытқулар саны.

Тәжірибеде дәлелденіп және ғылыми зерттеулердің нәтижелері бойынша расталып отыратын жайт – мәтін ағымын өңдеу барысында бұрмалану мен қателерді толығымен жоюдың мүмкін еместігі. Сондықтан, ең дұрысы, өңделу жұмысының дәрежесі жайлы ғана сөз ету (бұрмалану дәрежесі немесе қате

кету дәрежесі). Осындай жағдайды ескере отырып, бұрмалану дәрежесі мен мәтін түріне қарай әдебиеттерде тиісті талаптар қойылады.

Осы айтылғандарға байланысты ең жиі қолданылатын машиналық корректура әдістерін атап өтейік. Оларды машиналық қайтақұру және машиналық бақылау әдістері деп екіге ажыратуға болады. Машиналық бақылау әдістері, іштей тағы да түпнұсқаны қолдануға негізделген және белгіленген нормаға сай келетін символдар тізбегіне негізделген деп екіге ажыратылады.

Шығарманың түпнұсқа мәтінін пайдалануға негізделген әдістер ішінен ең кең қолданыс табатыны — қайталау (дублирование) әдісі. Оның мәні мынада. Мәтінді ақпаратты теру құралы арқылы таспаға (машина тетігіне) екі рет түсіріп алып, олардағы әрбір символды компьютерлік программалар көмегімен бір-бірімен салыстыру қажет. Егер қайталанған мәтіндер ішінде бір ғана позиция бойынша екі түрлі символ кездессе, онда қайталанған мәтіндердің біреуі бұрмаланған, яғни сол позицияда қате бар деген сөз. Мұндай бақылау әдісінің нәтижелілігі маман-корректордың іс-әрекетімен салыстырғанда жоғары және ол — 0,96-ға тең.

Нормаланған символдар тізбегіне негізделген бақылау әдістерінің ішінен ең көп тарағаны — сөздік әдісі. Оның мәні мынада: қайсыбір тілдің орфографиялық сөздігі ЭЕМ-ның есте сақтайтын сыртқы тетігіне жазылады да, түзетуге қажетті мәтіннің әрбір сөзі эталон ретінде алынып отырған ЭЕМ жадындағы сөздік бірліктерімен сәйкестендіріледі. Егер іздеген сөз сөздіктен табылса, онда ол сөз дұрыс жазылды деп саналады да, ал сөздік бойынан табылмаса — потенциалды дәрежеде (барынша) бұрмаланған сөз деп есептеледі. Мұндай бақылау әдісінің нәтижелілігі қайталау әдісіне қарағанда төмен дәрежеде және ол 0,5 пен 0,7 аралығындағы шамаларға тең.

Бұл жерде мынадай сұрақ туындауы мүмкін: норма бойынша (орфографиялық сөздік арқылы) мәтін көшірмесінің түпнұсқадан ауытқу жағдайын қалайша бақылауға болады? Мұндай мүмкіндіктің орын алуы, көпшілік жағдайда, көшірме мәтінінің түпнұсқа мәтінінен ауытқуы нормадан да ауытқу болып саналуына байланысты. Қазіргі кездегі баспа ісінің

тәжірибесі бойынша, қайтақұру (реконструкция) әрекеті әр уақытта автоматты түрде жүзеге асады деп саналады. Яғни мәтіндегі бұрмалауларды автоматты түрде іздеп тауып, оларды адам қолымен түзетуге болады деп ұйғарым жасалады. Солай бола тұра, формалданған мәтіндер үшін қайтақұру әрекетінің нәтижелілігін 0,8 бен 0,9 аралығында жүзеге асыруға болатын әдістер де бар.

Компьютер арқылы редакциялауда бақылау нормаларының саны редакциялау жүйесінің «редакциялау қуаттылығын» білдіреді. Бұл қуаттылық, айтарлықтай дәрежеде, шығарма мәтіні элементтерінің осы мақсатқа қаншалықты сай таңбалануына да байланысты болады.

Лингвистикалық нормалар, негізінен, жәй (қарапайым) мәтіндерді тексеруге мүмкіндік жасайды. Мұндай бақылау әрбір сөзге әртүрлі ақпараттар (морфологиялық, синтаксистік, прагматикалық және т.б.) қоса тіркеліп берілген сөздікке негізделеді. Лингвистикалық әдістердің ең қарапайым түрлері сөздердің кейбір критерийлері (өлшемдері) бойынша таңдалып алынған қайсыбір жиынды пайдалануға немесе, керісінше, кейбір сөздерді пайдаланбауға мүмкіндік тудырады. Күрделіленген әдістер сөздер арасындағы грамматикалық байланыстың, тыныс белгілерінің, мәтінді абзацтарға бөлудің, қысқартылған сөздердің және сөз тіркестерінің берілулерінің және т.б. дұрыстығын бақылауды жүзеге асырады.

Психо-лингвистикалық нормалардың да барынша кең спектрі қолданыс табуда (Мучник, 1985). Бұл нормалар белгілі оқырман қауымына арналады және сөз, сөйлем немесе фразадан тыс бірліктердің мүмкін болатын максимум «ұзындығын» белгілейді (бұларды *тез оқырлық* сипаттамалар деп те атайды). Осымен бірге, кейбір әдебиеттердің түрлеріне қарай, арнайы нормалар түрлері де белгіленеді. Мысалы, шет тілдерінің оқулықтары үшін мына жайттар тексеріледі:

- мәтін жаңалығы, яғни мәтіндегі оқушыға таныс емес (жаңа) сөздер санының мәтіндегі барлық сөздер санына қатынасы (нормалық мәні – 0,015 пен 0,50 аралығында);

- мәтіннің адаптациялануы (бейімделуі), яғни мәтіндегі бұрыннан да таныс сөздер санының оның жалпы сөздік қорына қатынасы (нормалық мәні – 0,237 мен 0,253 аралығында);

– баяндаудың қиындық сипаты (напряжение изложения), яғни оқушыға таныс емес сөздер санының оның жаңадан меңгерген сөздерінің санына қатынасы.

Лингвистикалық жабдықтау жүйесін редакциялық-баспа қызметін автоматтандыруға да қолдануға болады. Мәселен, мұндай мақсат үшін әртүрлі арнайы ақпаратты-ізвестіру тілдері құрастырылады және олар арқылы баспа мекемесіне жолданған хаттар (корреспонденциялар) өңделеді.

Құрастыру кезеңі. Бұл кезеңде басылымды құрастыру мақсатында тілдердің екі түрі қолданылады:

- а) полиграфиялық безендірудің бұйрықтар тілі;
- ә) мәтін элементтерін бөліп алуға қажетті таңбалар тілі (язык меток).

Бұл тілдер әртүрлі ұстанымдарға негізделіп жасалады. Жүйелерде таңбалар мен бұйрықтар саны, көбінде, 20 мен 100 сандарының аралығында орналасқан.

Көпшілік жағдайда полиграфиялық бұйрықтар мынадай құрылымға ие болып келеді:

- а) бұйрық белгісі;
- ә) бұйрық түрі;
- б) бұйрық параметрі (өлшемі).

Мәтін элементтері таңбасының (меткасының) құрылымы, негізінен, полиграфиялық бұйрық құрылымымен бірдей келеді. Бірқатар жүйелерде мәтін элементтеріне таңбалар тілі (язык меток) қолданылмайды да, ал оның орнына мәтін элементінің алдына полиграфиялық безендіру бұйрығы қойылады. Бұйрықтар мен таңбаларды (меткаларды) мәтінді полиграфиялық қайта туындату кезеңінен бұрынғы кезеңдердің барлығында да бірдей қоя беруге болады.

Полиграфиялық қайта туындату кезеңі. Бұл кезеңде баспа ісін автоматтандырудың лингвистикалық жабдықтау мүмкіндіктерінің ішінен, берілген форматтағы жолды қоспау үшін, тек бір ғана функция пайдаланылады. Ол функция сөзді тасымалдау қажеттігі үшін әрекет ететін компьютерлік программа түрінде көрініс табады. Мұндай программалар полиграфиялық безендіру бұйрықтарын өтеу үшін арналған мәтіндерді өңдеудің автоматтанған жүйелерінде қолданыстауып жүр.

Сөздерді тасымалдау программалары баспа ісін автоматтандырудың алдыңғы кезеңдерінде қызмет ететін басқа да жүйелерде қолданылуы мүмкін. Мысалы, мәтіндік процессорлар мен редакциялық-баспа жүйелерінде қолданылады. Бірақ бұл жүйелерде сөзді тасымалдау программасына қойылатын талап полиграфиялық қайта туындату кезеңіне қарағанда айтарлықтай қатал емес.

Сөзді тасымалдау алгоритмін екі үлкен топқа бөліп қарастыруға болады:

- сөздіктер арқылы, яғни бұл жағдайда компьютер жадында қайсыбір тілдің тасымалданған сөздерінің эталондық нұсқасы сақталады. Мұндай сөздердің көлемі 20-40 мың және бірнеше тілдер түрі (көбінде, оннан аспайтын тілдер) қамтылады;

- сөздіксіз, яғни бұл жағдайда тасымалдың белгісі тасымалдау ережесі бойынша компьютерлік программа көмегімен жүзеге асады.

Әрине, сөздіктік әдіс бойынша тасымалдау ісі аса нәтижелі болғанымен, мұнда әрекет ету мезгілі мен компьютер жадын үнемдеуден көптеп ұтылады. Ал сөздіксіз әдіс уақыт пен жадыға байланысты тиімді болғанымен, тасымалдау әрекетінің сапасы өте төмен.

Сөздіксіз әдістер ішінен мынадай кішігірім топтарды бөліп қарастыруға болады:

- морфемдік әдіс арқылы, яғни тасымалдау белгісін сөздің морфемдік құрамына қарай қою;

- морфемсіз әдіс арқылы, яғни тасымалдау белгісін буындар шекарасына қарай немесе құрамында дауысты және дауыссыз әріптердің екіәріптік тіркестерінің (биграммалардың) болуына қарай қою.

- қысындастыру әдісі. Бұл әдіс бойынша тасымалдау белгісі сөзформа құрамында префикстердің, суффикстердің және буындардың немесе биграммалардың болуына қарай қойылады.

Сөздіксіз әдістер арқылы жүргізілген зерттеулер нәтижесі бойынша, бұрыс қойылған тасымалдар саны, кері функционалдық байланыстың әсерінен қойылмай кеткен (жоғалған) тасымалдар санына қатысты болып келеді. Мысалы, егер қайсыбір алгоритм бойынша бұрыс қойылған тасымалдардың саны 1,12 пайызға тең болса, онда қойылмай кеткен (жоғалған)

тасымалдардың саны 48,41 пайыз құрайды. Сондай-ақ, керісінше, басқа бір алгоритм бойынша қойылмай кеткен (жоғалған) тасымалдардың саны 7,40 пайыз болса, бұрыс тасымалдардың саны 11,70 пайызға дейін өседі (*Будаар және т.б., 1985*).

Тарату кезеңі. Баспа ісін автоматтандырудың бұл кезеңі бойынша арнайы тасымалдау (немесе байланыс торабы) құралдары арқылы даяр машиналық ақпарат, яғни басылымдар нәтижесі тұтынушылар қауымына таратылады. Басылымдар – мерзімсіз немесе мерзімді басылымдар болып екіге ажыратылады. Мысалы, мерзімді басылымдарға «электронды» түрдегі журналдар мен газеттерді жатқызуға болады. Мұндай «электронды» басылымдар түрлері бірқатар Батыс Еуропа, Солтүстік Америка елдерінде және Жапония мен тағы да басқа елдерде дәстүрге айналып отыр.

Баспа ісін автоматтандырудың бұл кезеңінде мерзімді басылымдардың әр бөлігінен тұтас бір басылым құрастыру үшін, негізінен, лингвистикалық жабдықтау мүмкіндіктерінің әртүрлі рубрикаторлардың иерархиялық бір деңгейінде техникалық себептерге байланысты тек 10 рубрика ғана кездесіп жүр.

Баспа ісін автоматтандыруды лингвистикалық жабдықтау жүйесінің құрылымы. Мұндай құрылым автономды және жүйелі деп екіге бөлінеді. Лингвистикалық жабдықтаудың автономды түрінде түсіндірме және энциклопедиялық сөздіктер қамтылады. Жүйелі түрдегі лингвистикалық жабдықтауда деректердің тірек базасы, нормалардың жиыны және компьютерлік программалар сақталады.

Деректердің тірек базасының құрамы мынадай ақпараттардан тұрады: антонимдік, синонимдік, орфографиялық сөздіктерден, морфологиялық, синтаксистік, семантикалық және прагматикалық ақпараттар сөздіктерінен (жалпытілдік және салалық), эталондық тасымал сөздіктерінен (әртүрлі тілдер үшін), морфемалар тізімінен, өлшем бірліктері қысқартуларының тізімінен, персоналий тізімінен, полиграфиялық бұйрықтар тізімінен, мәтін элементтері таңбаларының тізімінен, библиографиялық сипаттаудың қалыпты элементтерінің тізімінен және арнаулы ақпараттық-іздістіру тілдері рубрикаторларының тізімдерінен.

Нормалар жиыны іштей мынадай лингвистикалық нормалар жиындарының деңгейі бойынша бөлінеді: графемдер, морфемдер, фразадан тыс бірліктер, блоктар (топтамалар), дискурстер; жоғары дайындық деңгейіндегі оқырмандар үшін психолінгвистикалық нормалар жиыны, төменгі дайындық деңгейіндегілерге – тасымалдау ережелерінің жиыны (өртүрлі тілдер үшін).

Компьютерлік программалар топтамасы мынадай құрамнан тұрады: авторлық және редакциялық кезеңдеріне арналған программалардан, құрастыру мен полиграфиялық қайта туындату кезеңдері үшін жазылған программалардан және тарату кезеңіне арнайы жазылған программалардан.

Лингвистикалық жабдықтаудың құрастырушылары бөлек және бірге көрініс табады. Мысалы, лингвистикалық ақпаратты сақтайтын сөздік құрамында орфографиялық және синонимдік сөздіктер де бірге берілуі мүмкін. Міне, осыған қарамай, лингвистикалық жабдықтаудың әрбір компонентін жеке бір нысан ретінде қарастыруға болады.

Соңғы жылдары дамыған дербес компьютерлер мен лазерлік принтерлердің негізінде *үстелүстілік баспа жүйесі* деп аталатын жүйелер құрыла бастады. Желі режимі бойынша да әрекет ете алатын мұндай үстелүстілік баспа жүйелер өзіне жүктелген қызметті қарастырылған кезеңдердің барлығында бірдей (тарату кезеңін есепке алмағанда) атқара алады.

Сонымен, лингвистикалық жабдықтау жүйесіне қойылатын негізгі талап – тек бір ғана лингвистикалық сөздік бойынша әрі синоним сөздерді іздестіруді, әрі корректураны, әрі редакциялауды, әрі сөздерді тасымалдауды жүзеге асыру.

Міне, осы аталған мәселелерді компьютерлік лингвистика тұрғысынан зерттеу баспа ісін автоматтандыру әрекетін жаңа сипаттағы жоғары деңгейге көтерері сөзсіз.

Әдебиет

Будаар Я. И., Карская С. Д., Партыко З. В. Сравнительный анализ алгоритмов переносов слов русского языка // Научно-техническая информация. Сер. 2. 1985. № 12. С. 20-30.

Мартин Дж. Видеотекст и информационное обслуживание общества. М., 1987. 184 с.

Мучник В. С. Человек и текст: Основы культуры письменной речи. М., 1985. 256 с.

Партыко З. В. 1) Методы машинной корректуры и машинного редактирования. М., 1983. 41 с. (Изд. дело: Обзорн. информ. / НИЦ «Информпечать». Вып.5); 2) Статистика ошибок при корректуре и редактировании текстов. М., 1989. 56 с. (Изд. дело: Обзорн. информ. / НИЦ «Информпечать». Вып.3); Лингвистическое обеспечение автоматизированного издательского процесса / Прикладное языкознание: Учебник, отв. ред. А.С. Герд. СПб: Изд-во С. –Петербург. С. 399-407...

Қазақ тілінің компьютерлік қорын (базасын) құрастыру жайлы

Кеңес Одағы кезіндегі кейбір ғажайып бастамалар заман өзгерісіне қарай құбылып, шешімі кезек күттірмейтін көптеген мәселелерге көзқарас басқаша қалыптасып, әрі қарайғы жалғасы табылмай қалғаны қазіргі кезде құпия бола қоймас. Солардың бірі – түркі тілдерінің, оның ішінде қазақ тілінің машиналық (компьютерлік) қорын жасау мәселесі.

Осы айтылған мәселенің қысқаша тарихы мен мән-жайына тоқталайық.

Электронды есептеу машиналарының (ЭЕМ) көмегімен ұлттық тілдерді зерттеу мен одан туындайтын ақпараттарды автоматты түрде танып-білудің негізінде XX ғасырдың 80-жылдарынан бастап тілдердің машиналық қорын (базасын) жасау жұмысы жайлы ой-пікірлер жүзеге аса бастаған болатын. Осыған байланысты 1983 жылы Мәскеуде шақырылған Бүкілодақтық ғылыми конференцияның күн тәртібіне орыс тілінің машиналық қорын жасау проблемасы алғаш рет қойылды. Бұл ғылыми мәжілісте ЭЕМ-ның жадына түсірілетін тілдік қордың құрамы мен құрылымы қандай болуы тиіс екендігі, бірінші кезекте мұндай қор қандай міндеттер атқаратыны, тіл көздерінің ауқымы қандай болмақтығы жайлы және т.б. көптеген мәселелер қаралған болатын. Сонымен бірге электрондық тіл қорын дәйекті түрде тұтыну үшін қандай ұйымдық және техникалық шаралардың атқарылу қажеттігі де ортаға салынып, жан-жақты талқыланды.

Аталған ғылыми конференцияның шешімдерінің орындалу жағдайын қадағалау мен жаңадан туындайтын мәселелерді шешу қажеттігінен 1987 жылы Мәскеуде ашылған Бүкілодақтық екінші ғылыми конференцияда осы саладағы ғалымдар тағы да бас қосты. Бұл ғылыми форумға Қазақстанның және басқа да ұлттық республикалардың ғалымдары қатысты.

Мәжілісте өткен төрт жыл аралығында орындалған жұмыстардың нәтижелерінің сөз болуымен қатар, басқа да ұлт тілдерінің машиналық қорларын жасау мүмкіншіліктеріне байланысты соны мәселелер қарастырылды. Ерекше атап кетерлік жәйт — осы ғылыми жиында орыс, украин, грузин, эстон тілдеріне қатысты баяндамалардың қатарында қазақ тілінің машиналық қорын жасау мәселесі де өз алдына тыңдалды.

Сөйтіп, орыс тілінің машиналық қорын жасау мәселесіне арналған екінші ғылыми конференцияның шешімдерінде қазақ мәтіндерін ЭЕМ жадына енгізу және оны автоматты жолмен статистикалық тәсілдермен зерттеу бағытындағы қазақ ғалымдарының қол жеткізген нәтижелері ескеріліп, түркі тілдерінің машиналық қорын жасайтын координациялық орталық Қазақ ССР ҒА Тіл білімі институтының құрамынан ашылсын деген дербес қаулы қабылданғаны біз үшін үлкен абырой болды.

Көп ұзамай, 1988 жылдың мамыр айында Мәскеуде Кеңес Одағы түркітанушы комитетінің XIV Пленумы ашылып, оның күн тәртібіне түркі тілдерінің машиналық қорын жасау мәселесі арнайы қойылды. Бұл ғылыми жиында негізгі баяндамашылар ретінде Одақ бойынша белгілі ғалымдар: Р.Г.Пиотровский, А.М.Щербак, В.Г.Гузев өз ой-пікірлерін ортаға салып, ондай қор жасаудың қажеттігімен қатар алда тұрған қиыншылықтарын да атап өтті (*Гузев, Пиотровский, Щербак*). Ал қазақ тілінің машиналық қорын жасау мүмкіндігі жайлы мәселеге байланысты қазақстандық ғалымдар баяндама жасады (*Қ.Б.Бектаев, А.Қ.Жұбанов*).

Тілдің машиналық қорының қажеттілігін алғаш дәлелдеп берген ғалым — информатика саласының белгілі маманы А.П.Ершов тілдің машиналық қорын құрастыру мәселесіне үлкен мән берген болатын (*Ершов, 1979*). Оған ғалымның 1978

жылы айтқан мына сөздері айғақ: «Орыс тілін сүйетін қоғамның (формалды қоғам) танымынша, орыс тіл білімі дегеніміз әзірге ұйықтап жатқан Гулливер тәрізді, сондықтан менің ойымша, оның оянатын және өзі туралы бар дауыспен жар салатын кезі енді жетті. Ал ЭЕМ бұл ретте таптырмайтын қоңыраулы сағат секілді» (*Машинный фонд...*, 7-12-б.). Бұл айтылғандар қазақ тілінің қазіргі жағдайының компьютерлік лингвистика саласына қатысты әлі де өз күшін жоғалтпаған тәрізді. Ал А.П.Ершов сол кездің өзінде-ақ аса білгірлікпен болжап, болашақ жасалатын орыс тілінің машиналық қорын тіл жөніндегі білімдердің аса қуатты концентраты (қоспасы) болады деп сенген.

Мұндай формалды компьютерлік жүйе тілдің жаратылысына сәйкес әрі тең көлемді болуы қажет, сөйте тұра ондай тілдік модель «анатомиялық» тұрғыдан мүшеленген, ашық, бақылауға айқын, зерттеуге қолайлы әрі өзгермелі болуы тиіс деп ұйғарамыз. Алайда қазақ тілін модельдеуге қатысты біздің жетістіктерімізді үндіеуропа тіл білімінің статусымен және оның компьютерлік лингвистика бойынша алған нәтиже ауқымымен салыстырып қарайтын болсақ, біз өтпек жолымыздың әлі де бастауында тұрғанымыз айқын сезіледі.

Көптеген ғалымдардың пікірлерін қорыта келе, біз «Қазақ тілінің компьютерлік қоры» атты жүйенің түсінігі — көп тармақты әрі терең автоматтандырылған жүйені аңғартады демекпіз және оның атқаратын міндеті ақпараттық барлау мен тіл білімі саласындағы (жалпы филология бағыттағы) зерттеушілік болуы тиіс. Ал ондай компьютерлік қор жасаудың методологиялық негіздері ретінде мынадай екі тезисті басшылыққа алғанды жөн санар едік. Оның біріншісі — тіл жөніндегі деректердің қандайы болмасын лексикографиялық пішінге келтірілуі қажеттігі, ал екіншісі — лексикографиялық жағынан жүйеленген тілдік мәліметтердің барлығын да компьютерлік алгоритмдер тіліне лайықтаудың керектігі.

Тілдік бірліктердің (ақпараттардың) лексикографиялық пішінде көрініс табуы олардың сөздік түріндегі жұмыстардың идеясына жақындайды, ал мұндай тұжырым тілдік ақпараттарды шоғырландыратын бірегей түйін немесе түп қазық — жеке сөз, деп ұйғарумен бірдей. Шынында да, сөз өзінің ішкі-сыртқы қызметтері жағынан және ақпарат сақтағыштық қасиеттерінің

негізінде белгілі бір жүйедегі сөз қорын танытатыны сөзсіз. Сондықтан сөздің белгілі мөлшерде өзіндік сегменттері болады және олар грамматикалық, стилистикалық, фонетикалық, контекстік т.б. пішінде көрініс табады, яғни ол барлық тілдік мәліметтер жүйесін сәулелендіреді деп ұйғарған жөн.

Егер біз ана тіліміздің компьютерлік қорын құру методологиясына орыс тілінің машиналық қорын жасау тәжірибесіне сүйенсек, онда мынадай қор құрамын сақтауымыз қажет:

- 1) қазақ тілі компьютерлік қорының негізгі сөзтізбегі (генеральный словник);
- 2) қазақ мәтіндерін суреттеу қоры (иллюстративный фонд);
- 3) қазақ тілінің терминологиялық қоры;
- 4) қазақ тілінің академиялық сөздіктері мен грамматика қоры;
- 5) қазақ тілінің лингво-статистикалық қоры;
- 6) қазақ тілінің лексикографиялық қоры;
- 7) қазақ тілі процессорларының қоры;
- 8) қазақ тіліне қатысты лингвистикалық алгоритмдер мен комплексті сипаттағы теориялық және тәжірибелік мәліметтерді бірлікте қамтитын бағдарламалар қоры.

Аталған құрамдағы қазақ тілінің компьютерлік қоры тілдік мәліметтердің сала-салаға жүйеленген банк ретінде де және тұтыну қажеттігіне сай қайта жинақталған тұтас күйінде де пайдалану мүмкіндігі алдын ала ескерілуі керек.

Қазақ тілі жөніндегі деректерді жинаудың, сақтаудың, талдаудың және салыстыра зерттеудің жаңа әдістеріне көшу, сондай-ақ лингвистикалық дерек көздерінің автоматтандырылған сөздіктер мен грамматикалар тәрізді жаңа түрлерін игеру ісі өмірге жанасымды және тиімді екені сөзсіз. Әрине, бұл үшін біз жалпыфилологиялық дәстүрлер мен мәдениетке, тілді танып білуге, практикалық жаңа міндеттер тәжірибелерін мұқият ескеруге тиіспіз.

Болашақ қазақ тілінің компьютерлік қорының құрамындағы «қазақ мәтіндерін суреттеу қорының» мақсаттары мен міндеттері қандай болмақ?

Біздің ұйғаруымызша, мұндай жүйелерде тұтас қалыптағы мәтіндер сақталуы қажет те және тұтынушының сұранысы бойынша сол мәтіндердің үзінділері, қор ішінде тұрған олардың

нақты орны, сөзтұлғалардың статистикасы, сөзнұсқағыштардың толық тізімдері, сөздерге телулі цитаталар, жиілікті және кері әліпбилі сөздіктер көрініс табуы қажет.

Компьютерлік қордың лексикалық сипаты түпнұсқа мәтіндерінің типіне байланысты деп санай отыра, қазақ тілінің жүйесін танытатын сөйленіс жағдаятына: көркем әдебиет тілін, газет-журнал тілін, ғылым мен техника тілін, ұйымдастыру-нұсқау құжаттары тілін, заң орындары мен сот мекемелер тілін, оқу орындары тілін, сауда, дәріхана, емханалардағы және т.б. сөйленіс түрлерін жатқызуды қолдар едік. Әрине, қордың бұл түріндегі компьютер жадына орналастыруға қиындық туғызатын сөйленіс түрлеріне жататындар: диспуттар, сот процестері, жиналыстар, күнделікті әңгімелесулер, телефон арқылы сөйлесу т.т. сияқты ауызекі сөйлесу тілінің үлгілерін жазып алу мәселесі.

Келесі бір қомақты мәтіндік бірлік — ол терминдер мәселесі. Қазіргі ғылыми-техникалық және ісқағаздары мәтінінің негізгі өзегі терминдер екені айқын. Жалпы сөйлеу тілінің аясындағы сөздер емес, терминдердің тілдік қоршауы (мәнмәтіні) өзгеше бір тілдік жүйе, басқаша айтқанда, терминдер өрісі болып табылады. Осы аяда өмір сүретін термин үшін барлық ғылыми шығармалардың сөз қолданыс машықтары етене, ортақ болып келеді.

Осындай маңызды жұмыстардың келесі түрі — әр дәуірдегі мәтіндердің тілдік белгілерін ажыратып ала білу мәселесі. Осы мақсатта тіліміздегі диалектілік лексиканың біркелкі еместігін ескеретін болсақ, оларды компьютерлік қорға қосу жұмыстарының қазақ диалектологтары үшін ойланатын жақтары жоқ емес. Мәселен, олар бір жағынан ертедегі сөздер мен қолданыстардың ұлтқа, жазуға дейінгі үлгілері болып келсе, екінші жағынан басқа тілдерден енген сөздер, таралу өрісі шектеліп қалған сөздер немесе тіліміздегі жаңадан енген сөздер болуы да ықтимал.

Халық тілі сол халықтың жанды қазынасы болғандықтан, соның бүкіл болмысын барлық ерекшеліктерімен қоса қамту болашақ «Қазақ тілінің компьютерлік қорының» міндеті деп түсіну қажет сияқты.

Жоғарыда сөз болған қазақ тіліндегі диалектілер секілді, әрі синхрониялық, әрі диахрониялық сипатта күн кешетін тіл —

фольклордың тілі және жер-су аттарының тілі немесе топонимика.

Осы аталған тіл үлгілерін жинастырып, тіліміздің компьютерлік қорын жасауда, оның бұл аталғандардан басқа да салалары мен арналары қамтылары сөзсіз.

Формалды пішіндегі қазақ тілінің компьютерлік қорын жасау мақсатымен байланысты күтілетін ең маңызды мәселелер мыналар деп білеміз:

а) тұтынушыға қажет болған белгілі бір сөзді қандай тіл көздерінен, мәтіндерден, тіл айырымдарынан іздестіру қажет;

ә) іздеуге тиісті сөз қарастыратын мәтінде қамтылған ба, жоқ па?

б) графикалық және фонетикалық пішіндерін қоса есептегенде сөздің мүмкін болған барлық түрлері мен тұлғасы айқындалып болды ма?

в) белгілі бір сөзге қатысты мәтін үзінділерінің (мәнмәтіндердің) жалпы аумағы, сипаты анықталды ма?

г) белгілі бір сөз: қарастырып отырған семантикалық және грамматикалық сипатта өткен дәуірлердегі, көне ғасырдағы мәтіндерде қолданылған ба, егер қолданылса, қандай сөздіктерден көрініс тапқан?

Міне, осы аталғандар қазақ тілінің компьютерлік қорын жасау мақсатындағы күн тәртібіне қойылатын ең маңызды мәселелер болуы қажет.

Жоғарыда сөз болған Совет түркологтары комитетінің XIV Пленумына қайта оралып, мәжілісте көтерілген негізгі мәселелерге тоқтала кетейік.

Алғашында мәжіліске қатысушылар, тілдің машиналық қорының анықтамасы төңірегінде пікір алмасып, соның нәтижесінде бір түйінге келді деуге болады.

ЭЕМ-ның (компьютердің) араласуымен зерттелетін жекеленген мәселелердің нәтижелерін, мысалы, жеке тұрып лексика-семантикалық, этимологиялық, морфологиялық, тіпті, синтаксистік есептерді шешетін машиналық бағдарламалардың жиынтығын немесе ЭЕМ-ның жадына енгізілген мәтіндер мен лексикографиялық материалдардың тұтастығын сол тілдің машиналық қоры, міне осы, деп есептеуге болмайтындығы баса көрсетілді.

Мәселен, ҚР БЖҒМ А.Байтұрсынұлы атындағы Тіл білімі институтында компьютерлік лингвистика саласы бойынша біршама ғылыми жұмыстар орындалды. Оларды қысқаша былайша атап өтуге болар еді:

- көркем әдебиет, көсемсөз, ғылыми-көпшілік әдебиет пен оқулықтар мәтіндерінің, алғашында ЭЕМ-ның, кейіннен компьютердің жадына көптеп енгізіліп, олар бойынша арнайы компьютерлік бағдарламалар құрастырылуының негізінде әртүрлі жиілік сөздіктердің алынуы;

- М.Әуезовтің 20 томдық шығармалар жинағының мәтіндерінің компьютер жадына енгізілуі және соның негізінде жазушы тілінің жиілік сөздіктерінің жарық көруі;

- осы тәрізді жиілік сөздіктердің негізінде қазақ сөздерінің әр стильде және әр көлемді мәтіндерде қайталануының қосынды жиілігінің, сол мәтіндерді қамту статистикасының анықталуы;

- сөзтұлғалардың құрылымдық статистикасы мен мәтіндегі пайыздық салмақтарының әр сөз таптарына қатысты айқындалуы;

- қазақ тіліндегі негізгі сөз таптарының мәтін бойында кездесулерінің математикалық статистика мен ықтималдық теория заңдылықтарына бағыну-бағынбау жағдайларының анықталуы;

- қазақ тілінің «ТІЛ – ҚАЗЫНА» атты автоматтандырылған компьютерлік картотекалық қорының іске қосылуы;

- қазақ тіліндегі мәтін мазмұнын оның тұрпатына қарай ашудың (тұрпаттаудың, формалдаудың) негізгі принциптерінің айқындалуы;

- Тіл білімі институтында жоспарлы тақырыптар бойынша түзіліп жатқан түсіндірме, екі тілдік, бір тілдік сөздіктер мәтіндерінің компьютер жадында «сөздіктер қорын» құрап, тұрақты сақталуы.

Міне, осы жұмыстар түрлерінің жиынтығын біз қазақ тілінің компьютерлік қоры деп атауымызға бола ма?

Жоқ. Себебі, бұл аталғандар жеке мәселелерді шеше алатын ғана бағдарламалар мен мәтіндік файлдар жиынтығы және олар бір-біріне тығыз байланыста болмай, тұтынушы сұранысына бейімделмегені әрі бір орталықтан басқарылатын

комплекті түрдегі автоматтандырылған күрделі жүйе болмауынан деп білеміз.

Жоғарыда сөз болып жатқан Пленум мәжілісінде түркі тілдерінің көпаспектілі машиналық қорын жасау қолданбалы ілімдердің мұқтаждығына сай жүзеге асырылуы тиіс екендігі баса айтылды.

Өкінішке қарай, мұндай жақсы бастама сөз жүзінде ғана қалып, түркі тілдерінің көпаспектілі машиналық қорын өмірге келтіру әлі де болашақтың ісіне айналып отыр.

Ал қазіргі жағдайда, ең алдымен, егеменді еліміздің мемлекеттік тіліне айналған ана тіліміз — қазақ тілінің компьютерлік қорын ғылыми тұрғыда құру — жаңаша ойлап, жаңаша шешудің ең тиімді жолы деуге әбден-ақ болады.

Енді түркі тілдерінің көпаспектілі машиналық қорын құруға қажетті деп есептелетін түрліше салаларына (блоктарына) жалпылама сипаттама бере отыра, сол идеяны қазақ тілінің компьютерлік қорын жасау кезінде де пайдалануға болады-ау деген оймен, оның төменде аталатын тарамдарын оқырман қауымға ұсынып отырмыз.

1. Алдымен құрастыруды қажет ететін сала — лексикалық блок. Ол орфографиялық, орфоэпиялық, түсіндірме, терминологиялық, жиілік және аударма сөздіктерді қамтитын көпшілік қолды лексикографиялық жүйе болып табылады.

2. Грамматикалық блок — лексикалық және грамматикалық морфемалардың тізбелерінің жиынтығынан, сондай-ақ синтаксистік модельдерді (сызбалық) қамтитын тілдік материалдардан құралады.

3. Фонетикалық блок — тілдің фонемалық құрамын анықтайды, әр фонеманың айырым (дифференциал) белгілерінің анықтамалары беріледі.

4. Морфонологиялық блок:

а) тілдік бірліктердің фонемдік құрылысының ерекшеліктері жөніндегі деректер қамтылады. Мәселен, негіздер, форманттар, сөздердің жекелеген кластары (бейнелеуіш сөздер) т.б.;

ә) тілдің сингармонизм заңдылығы, сондай-ақ дауыссыз дыбыстардың ассимиляциясы, қатандауы мен ұяндауы жайындағы мәліметтер ескеріледі;

б) екпіннің орны көрсетілген түбір негіздер мен сөз-формалар жинақталады.

Міне, осы аталған салалардың бәрін жан-жақты қамти отыра, қажетті деген сұранысқа толық та және тұжырымды жауап алу үшін біз соңғы ғылыми-технологиялық жетістіктерді пайдалана отырып, әрі жүйелі, әрі сапалы түрде көрініс табатын ғылыми тұрғыдағы қазақ тілінің компьютерлік қорын іске қосуымыз қажет.

Ұсынылып отырған жоба іске асып жатқан жағдайда мынадай сұрақ туындауы мүмкін: «Қазақ тілінің компьютерлік қоры» жасалып болғаннан кейін ол тек архив тәрізді қызмет ете ме, әлде ол тіл жөніндегі жаңа білімдер қосатын, жаңа мәліметтер тауып беретін ғылыми зертхана тәрізді бомақ па?

Бұл ретте айтарымыз, екі жақты құбылыстың соңғысы алғашқы қисынның барлық қасиеттерін иемденетіндігі себепті, оның берер мүмкіншіліктері де молырақ болмақ. Демек, «Қазақ тілінің компьютерлік қоры» бізге белгілі тілдік заңдылықтар жөніндегі мәліметтерді бойына жинақтаған тілдің синхронды және диахронды процестерін модельдей алатын әрекет үстінде болуы қажет. Сонда ғана компьютерлік тілдік қор түрлі деңгейдегі тіл қабаттарымен жүйелі негізде қызмет ететін болады. Басқаша айтқанда, «Қазақ тілінің компьютерлік қоры» дегеніміз қазақ тілі жайлы барлық мәліметтерді жинақтаушы банк, бірақ ол «ақылды», мамандандырылған білім қоры. Оны өзімізге дәстүрлі кітап ретінде де пайдалануға болады және зерттеуші өз ұсыныстарын белгілі бір мақсаттарда жүзеге асыра алады деген сенімдеміз.

Осы саладағы белгілі ғалымдардың тұжырымдауынша, компьютерлік тілдік қор дегеніміз ғылым адамының өз зерттеу нысанына жаңаша тұрғыда көз салудың мүмкіндігі болып табылады. Мұндай тілдік қор неғұрлым қомақты болса, солғұрлым тіл құрылысының сырын тереңірек аша алады, сөйтіп зерттелетін нысан жөніндегі түсініктердің шеңбері кеңиді, адамның білім өрісіндегі «ақтандақтардың» бедер-бейнесі айқындала түседі.

Біздер де осы айтылғандарға қосыла келе, өз атымыздан «Қазақ тілінің компьютерлік қоры» зерттеуші адамның қалып-қабілетін әлденеше есе арттырады, шығармашылық қуат көздері

ашыла түседі, сөйтіп бұл жаңа мүмкіндіктер қазақ тілінің жүйелілік қасиеттерін жетілдіруге және тіл жүйесін мұқият тануға жұмсалар еді деп тұжырымдар едік. Әрине, бұл аталып отырған тілдік қорды өмірге келтіру үшін оның жоба-жолдары мемлекеттік дәрежеде, яғни ҚР Білім және ғылым министрлігі, Тіл комитеті деңгейінде қаралуы қажет. Себебі, мұндай жалпы филологиялық, ұлттық және мәдениет дәрежесіндегі күрделі жүйені іске қосу – қажетті мамандарды (жоғары дәрежелі инженер-бағдарламашыларды, компьютерлік лингвистика саласының мамандарын және т.б.) және тиісті қаражат мүмкіндігін шешуді талап етеді.

Сөз соңын, машиналық қор жайындағы идеяны осыдан ширек ғасыр бұрын ең алғаш рет ұсынған информатика саласының белгілі ғалымы А.П.Ершовтың сол кездердегі айтқан сөзімен қорытындылағымыз келіп отыр.

А.П.Ершов 1978 жылы жарық көрген «Іс қағаздары прозасының феномены: диалогтар жүйесін жасау методологиясына» атты мақаласында: «Егер орыс тілінің машиналық қорын жасау мақсаты ойдағыдай орындалса, модельдеу мен алгоритмдерді құрастыру саласындғы қол жеткен табыстарымыздың бәрі академиялық жаттығулар қалпында қалып қоймай, іс жүзіне асқан болар еді, мұндай өте-мөте қажетті проблеманың шешілуі ғылым, мәдениет және қолданбалы техникалық тәсілдер тұрғысынан мейлінше құнды болмақ. ...Тілдің машиналық қорын жасау ісі өндірістік лингвистикалық жүйелерді іске қосудан озық болуы тиіс, өйткені тек осы жағдайда ғана толып жатқан қажетсіз қайталаулардан құтылып, тілдің жанды өзегін озбырлар мен дүмшелердің қиянатынан аман сақтап қалуға болады» (Ершов, 1979).

Әдебиет

1. Гузев В.Г., Пиотровский Р.Г., Щербак А.М. О создании машинного фонда тюркских языков // Советская тюркология. 1988. №2. С. 98-101.
2. Ершов А.П. К методологии построения диалоговых систем: феномен деловой прозы. Препр. №156, ВЦ СО АН СССР. Новосибирск, 1979.
3. Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. 240 с.



Қ О Р Ы Т Ы Н Д Ы

Тілдік құбылыстардың біркелкі еместігі, бір қарағанда, компьютерді тіл білімі саласындағы зерттеу жұмыстарында пайдалануға кедергі келтіретін сияқты. Егер біз компьютердің мүмкіндігін толық түрде ескере білсек, тіпті шешімі жоқ-ау деген көптеген мәселелердің де шешімі табылар ма еді. Көбінде, компьютер жадында орындалатын амалдар тілшінің орындайтын формалды іс-әрекетіне ұқсас және олар көрініс деңгейінде тіл мен сөйлеу аспектілерінің көптеген жақтарын қамти алатыны сөзсіз. Сондықтан бұл жердегі ең бастысы адам баласы өзінің ойлау процесін компьютерге бөлшектеп беріп, формалдау жолдарымен ұштастыра білуі қажет сияқты. Сонда ғана зерттеуші тілдік парадигмалық көріністерді, фразалық тармақтандудың заңдылықтарын ашуды және т.б. формалдануға қажетті жұмыстарды іске асырар ма еді.

Қазақ тіл біліміндегі зерттеу кезеңдеріне, әсіресе, лингвистикалық ақпараттармен жұмыс істеу: оларды жинақтау, іздеу, сақтау, шет тілдегі ақпараттарды қазақ тіліне аудару, ақпараттық библиографиялық іздеу, қазақ тіліндегі мәтіннен реферат алу және т.б. істерге логикалық роботтың (компьютердің) араласуы жақын болашақта мемлекеттік тұрғыда қайта қаралып жатса, бұл мәселе өндірістік дәрежеге дейін көтерілуі де мүмкін.

Ғылым мен техниканың дамуы тіл қызметінің қоғамдағы коммуникативтік қолданыс аясын кеңейте түскені бәрімізге

мәлім. Тіл өзінің адамдар арасындағы қарым-қатынастық қызметін атқаруымен бірге адам мен компьютердің қазақ тіліндегі диалогын да жүзеге асыруы қажет деп білеміз. Осы себептен жақын болашақта табиғи тіл және оның түрлері ақпарат таратудағы «адам—компьютер—адам» аралығындағы коммуникативтік жүйе негізін құрайды деп сенім білдіруге әбден болады. Әрине, аталған жүйеде – тіл, өзінің адам баласының ойлау қабілетіндегі үйреншікті ассоциативті-эвристикалық мүмкіншілігінен басқаша, өзіне жат қалыпқа түсіп, компьютердің тізбектік-алгоритмдік субстратына ауысады. Тілдің бұл құбылысы қазақ тіл біліміндегі ізденістерге біршама методологиялық қиыншылықтар туғызуы мүмкін. Осы себептен, қазіргі тілдік зерттеулердің, біріншіден, қатаң логикалық теорияға бағындырылуын, екіншіден, объективті және толық болуын барынша талап ету қажет.

Тіл қызметінің жоғарыда сөз болған шарттары тіл білімінің «инженерлік лингвистика» деп аталатын жаңа саласында қарастырылады.

Инженерлік тіл білімі, негізінен, электрондық машина (компьютер) қажеттігі үшін тілдік белгілерді, тіл жүйесі мен оның атқаратын қызметін (функциясын) модельдеу мәселелерімен айналысады.

Тілдік объектілерді жасанды түрде қайта туындату үшін, тек қана жоғары дәрежедегі техникалық база болуы жеткіліксіз. Біздің ұйғаруымызша, қолданбалы есептер үшін, ең бастысы, күшті және объективті теория құрылуы қажет. Сондықтан да әрбір зерттеушінің тәжірибе жүзінде алған нәтижелері сол лингвистикалық теорияға баға беретіндей өлшемге айналуы керек.

Инженерлік тіл білімінің қажеттігін дәлелдейтін проблемалық бағыт-бағдарлардың ең негізгісі – мазмұн бірлігін автоматты түрде айқындап білу және табиғи тіл мен жасанды тілдерді семиотикалық жағынан барынша жақындастыру проблемалары. Сонымен қатар ашық жүйеге жататын табиғи тіл және келісіммен алынған, жабық семиотикалық жүйедегі қарым-қатынастарды зерттеу – инженерлік тіл білімінің методологиялық және технологиялық мәселелерінің ең бір негізгілері. Бұл мәселенің дұрыс шешім табуы – семиотика, информатика, компьютерлік

бағдарлама жасау және математикалық машиналар теориясы, ойлау процестерін модельдеу салаларындағы ғылымдардың қарқындап дамуының бірден-бір шарты.

Инженерлік лингвистиканың тіл біліміндегі басқа да пәндер қатарынан тең дәрежеде орын алуы — көптеген тілшілер мен математиктердің, инженер-программистердің жүргізген тәжірибелерінің, теориялық ізденістерінің нәтижесі деп түсінген жөн. Мәселен, қазіргі қазақ тілінің «Статистикалық лингвистика» және «Инженерлік лингвистика» салалары — физика-математика ғылымдарының кандидаты, филология ғылымдарының докторы, профессор Қ.Б.Бектаев есімімен және ол ғалымның шәкірттерінің көптеген ғылыми ізденістерімен тығыз байланысты. Сол сияқты, қазақ тілі грамматикасын формалды сипаттауға байланысты соңғы кездегі ғылыми ізденістерімен көзге түсіп жүрген техника ғылымдарының докторы, профессор А.А.Шәріпбаевты және т.б. атауға болар еді. Аталған зерттеу нәтижелері қазақ тілін инженерлік лингвистика әдіс-тәсілдерімен қарастыруда, мемлекеттік тілімізді оқытудың (үйретудің) автоматты түрдегі жүйесін құруда және қазақ тілінің формалды грамматикасының ғылыми негізін құру жолының бастамасы болатынына еш күмәніміз жоқ.

Классикалық лингвистика, көбінде, ақпараттық процестің қандай материалдық ортада жүріп жатқанына мән бере бермейді. Ал инженерлік тіл білімінде басқаша, яғни табиғи тілдің «адам—компьютер—адам» жүйесіндегі қызметінің ерекшеліктері стратегиялық және технологиялық ортаға тікелей қатысты. Сондықтан тілдік мәліметтерді компьютер арқылы қайта қарастыру жағдайында, адам баласының миы мен компьютер жадының құрылымды-функционалдық сипаттарын салыстыра отырып, оларға тән қасиеттерді ескерген жөн. Олай дейтініміз, тіл білімі саласындағы электрондық машиналар мен адам баласының мүмкіншіліктері жайлы көптеген тілшілер мен математиктер ғылыми әдебиеттер бетінде өз ойларын білдіріп жүрсе де, тіл білімі саласындағы компьютердің болашағы жайлы ойлары әлде де бірізді емес. Мысалы, біреулері ЭЕМ-ның жады мен жылдамдығын місе тұтпаса, ал басқалары — компьютер арқылы семантика-синтаксистік, тіпті стильдік ерекшеліктерге қатысты күрделі-күрделі мәселелердің болаты-

нын сөз етіп, қалай болғанда да компьютерді тілден алшақтатуға тырысатындар да жоқ емес.

Қазір, компьютерлік техниканың барынша дамыған кезінде, оның жадының көлемі, жылдамдығы туралы сөз болмаса да, басқа ғылыми мәселелердің «кезек күттірмейтін» тұстары (мысалы, түрлі сөздіктер түзу) себеп болып, қазақ тілін зерттеуде компьютерлік тіл білімінің жолы ойлағанымыздай болмай, қарқыны бұрынғыдан да бәсеңсігенге ұқсайды.

Адам мүмкіншілігіне қарағанда, компьютер, сырт кедергіге қарамай, жадындағы ақпаратты әрі тұрақты, әрі берік, әрі ұзақ мерзімге сақтайтыны және қажетті ақпаратты тез арада қағаз не монитор бетіне шығарып бере алатындығы көпшілікке мәлім. Ал мәтін ішіндегі ақпаратпен (мысалы, мағына мен мазмұн) жұмыс жасауда адамның ми қызметі оның физиологиялық және психикалық қалпына байланысты болатын болса, мұндай қасиет компьютерге жат деуге болады.

Инженерлік тіл білімінің даму болашағын сөз еткенде, мысалы, жақын он жылдың ішіндегі зерттеудің стратегиясын анықтауда және үйлесімді технологияны таңдауда, тек қана бүгінгі күнгі кибернетикалық техниканың және белгілі шеңбердегі тіл білімі мен нейрофизиологияның ғылыми жетістіктеріне сүйенген жеткіліксіз деп есептейміз. Бұл жерде кибернетикалық техниканың даму қарқынын ескере отырып, кибернетиканың, тіл білімінің және нейрофизиологияның методикалық жақтарын ескеретін кең шеңбердегі алдын болжай алатындай ғылыми көзқарас қажет.

Қазіргі кезде еуропа тілдеріндегі мәтіндік хабарламалардың грамматикалық, сөздік, фразеологиялық, кейбір лексика-грамматикалық және, тіпті, семантикалық деңгейлерін айқындайтын компьютерге арналып жасалған алгоритмдер негізіндегі бағдарламалары іске қосылған.

Осы айтылғандардың негізінде төмендегідей сұрақ туындауы мүмкін.

Электронды автоматтың (компьютердің) болашақтағы мәтінді түсіндіру мүмкіндігі қай деңгеймен шектелмек? Ол тек лексика-грамматикалық, лингвистикалық деңгей ме? Әлде компьютердің мәтін мазмұнын анықтаудағы мүмкіндігі маман аудармашының деңгейіне теңелуі мүмкін бе? Компьютерлік

аударма проблемасына қатысты бұл сұрақты былай да қоюға болады: шетел тіліндегі белгілі бір сала бойынша алынған мәтінді әрі осы саланың маманы, әрі тәжірибелі аудармашы болып табылатын адамның аударуы мен компьютерлік аударманы салыстыруға бола ма? Әлде компьютерлік аударма еш уақытта жоғары сапалы болмай, оған тұтынушының амалсыз араласуы қажет пе?

Әрине, қазіргі кездегі компьютерлік аудармалар әлде де маман аудармашының жәрдемін қажет етеді. Сондықтан да кибернетикалық автоматтың мәтіндерді «адамша» аударуы бір адамның бойына біткен лингвистикалық білім дәрежесімен және басқа да табиғи мүмкіншіліктерімен шектелмеу керек, ол компьютер үшін жеткіліксіз. Болашақтың лингвистикалық компьютері бүкіл бір ұйымға тән, әлеуметтік кәсіби (профессионал) топтарға тән интеллектуалдық, мәдени-әлеуметтік және лингвистикалық тәжірибелерді өзіне игеруі қажет.

Сонымен, хабардың берілу жүйесінің салалық тілдермен жұмыс істеуі, табиғи тіліміздің коммуникативтік қызметін дамыта түсетіндігіне еш күмән жоқ деуге болады. Ал қазақ тіл білімінің компьютерлік саласының тиісті дәрежеде дамуын, оның қатаң ережеге бағынуынан, мәтін бойындағы тілдік бірліктерді тануға қажетті формалды қасиеттерінің айқындалуынан, омоним сөздер мен көпмағыналы сөздердің және синоним сөздер қатарын сирету мен реттеу мәселелерінің өз шешімін дұрыс табуынан байқауға болар еді.



МАЗМҰНЫ

Алғы сөз	3
Кіріспе	7
«Адам—компьютер—адам» жүйесіндегі табиғи тіл қызметінің ерекшеліктері	20
Жасанды интеллект жүйесін құрастыруға қатысты мәселелер жайында	28
Оқу лексикографиясының ғылыми пән ретіндегі ерекшелігі	39
Ғылыми-техникалық лексикография және арнаулы сөздіктер түрлері	52
Лексикография саласындағы автоматтандыру тәсілі	64
Реферат құрастырудың әдіс-тәсілдері	84
Ақпарат іздестіру жүйесі	97
Автоматтанған оқу жүйесінің негізгі ұстанымдары	129
Автоматты (машиналық) аударма	140
Баспа ісін автоматтандырудың лингвистикалық негіздері	177
Қазақ тілінің компьютерлік қорын (базасын) құрастыру жайлы	189
Қорытынды	199

О қ у б а с ы л ы м ы

Жұбанов Асқар Құдайбергеноұлы

**КОМПЬЮТЕРЛІК ЛИНГВИСТИКАҒА
КІРІСПЕ**

О қ у к ұ р а л ы

ИБ № 3811

Басылуға 07.06.2007 жылы қол қойылды. Пішімі 60x84 1/16.
Көлемі 12,75 б.т. Офсетті қағаз, RISO басылыс. Тапсырыс № 100.
Таралымы 500 дана. Бағасы келісімді. Әл-Фараби атындағы
Қазақ ұлттық университетінің «Қазақ университеті» баспасы.
050038, Алматы қаласы, әл-Фараби даңғылы, 71.
«Қазақ университеті» баспаханасында басылды. *Қ*



Жұбанов Асқар Құдайбергенұлы - филология ғылымдарының докторы, профессор, А. Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері. Негізгі мамандығы - математик. Ұзақ жылдар бойы жоғары оқу орындарында математикадан дәріс берген. Әкесі - қазақ тілінің теориялық негізін қалаушы, қазақ филологиясының тұңғыш профессоры Құдайберген Қуанұлы Жұбанов. Әке ісін жалғастыру мақсатымен А.Қ.Жұбанов қазақ тілін математикалық әдістермен зерттеу

мәселесімен шұғылданады. Аспиранттық және докторанттық кезеңдерін Минск қаласының Лингвистикалық университетінде өткізіп, осы саланың белгілі ғалымдары - Р.Г.Пиотровский, А.В.Зубов, Қ.Бектаевтың ғылыми мектебінен өтеді. 1971 жылы "Статистико-лингвистическое исследование казахского текста с применением ЭВМ" атты тақырыпта кандидаттық диссертация қорғап, сол жылдан бері А.Байтұрсынұлы атындағы Тіл білімі институтында статистикалық, компьютерлік лингвистикаға және қазақ тілі деректерінің компьютерлік базасын құрастыруға қатысты тақырыптарға ғылыми жетекшілік етуде. 2001 жылы "Основные принципы формализации содержания казахского текста" деген тақырыпта докторлық диссертация қорғады. Аталған сала бойынша 300-ге жуық ғылыми мақала, 2 монография, 3 жиілік сөздік (авторлардың бірі) жарияланды. Соңғы 5-6 жыл бойы әл-Фараби атындағы Қазақ ұлттық университетінің және Абай атындағы Қазақ ұлттық педагогикалық университетінің филология факультеттерінің магистранттары мен студенттеріне математикалық және компьютерлік лингвистика пәндері бойынша дәріс беру нәтижесінде "Қолданбалы лингвистика: қазақ тілінің статистикасы", "Қолданбалы лингвистика: формалды модельдер" және "Компьютерлік лингвистикаға кіріспе" атты оқу құралдары жарық көрді.