

Г. Б. Мәдиева, С. Б. Бектемірова, Ж. Ж. Күзембекова

ЛИНГВИСТИКАЛЫҚ  
ЗЕРТТЕУЛЕРДЕГІ  
КОМПЬЮТЕРЛІК  
ТЕХНОЛОГИЯЛАР

*Оқу құралы*

Стереотиптік басылым

Алматы  
«Қазақ университеті»  
2020

ӘОЖ 80/81 (075)

КБЖ 81.2я7

М 13

*Баспаға әл-Фараби атындағы Қазақ ұлттық университеті  
филология, әдебиеттану және әлем тілдері факультетінің  
Ғылыми кеңесі және Редакциялық-баспа кеңесі  
шешімімен ұсынылған*

**Пікір жазғандар:**

филология ғылымдарының докторы **Б.Қ. Мұрзалина**  
филология ғылымдарының кандидаты **Ж.М. Уматова**

**Мәдиева Г.Б. және т.б.**

М 13 Лингвистикалық зерттеулердегі компьютерлік технологиялар: оқу құралы / Г.Б. Мәдиева, С.Б. Бектемірова, Ж.Ж. Күзембекова. – Стер. бас. – Алматы: Қазақ университеті, 2020. – 119 б.

**ISBN 978-601-04-0917-0**

Оқу құралында лингвистикалық зерттеулердегі компьютерлік технологиялардың негізгі түсініктемесі, ақпараттық технологияның лингвистикадағы қолданылу аясы, сонымен қатар компьютерлік лингвистиканың қолданбалы салалары және т.б. мағлұматтар беріледі.

Оқу құралы дәріс мәтіндері, семинар тапсырмалары, негізгі, қосымша әдебиеттер тізімінен және аралық бақылау бойынша тест тапсырмаларынан тұрады. Кітап соңында компьютерлік лингвистика ғылымы бойынша жарық көрген әдебиеттер көрсеткіші берілген.

Оқу құралы ғылыми жұртшылыққа, филологтарға, жалпы жоғары оқу орындарында білім алатын барша студенттерге, магистранттарға, PhD докторанттар мен ізденуші ғалымдарға, т.б. арналады.

**ӘОЖ 80/81(075)**

**КБЖ 81.2я7**

ISBN 978-601-04-0917-0

© Г. Мәдиева, С. Бектемірова, Ж. Күзембекова, 2020

© Әл-Фараби атындағы ҚазҰУ, 2020

## КІРІСПЕ

Бәсекеге қабілетті білім кеңістігін құруға бағытталған білім беру реформасының басты мақсаттарының бірі – инновациялық технологияларға сүйене отырып, болашақ маман даярлауда олардың біліктілігін дамытуға мүмкіндік жасау. Осындай күрделі міндеттерді шешуде ЖОО-да студенттерді жаңа инновациялық технологиялардың көмегімен оқытудың алатын орны ерекше. Инновациялық білім беру құралдарына: *аудио-видео құралдар, компьютер, интерактивті тақта, интернет, компьютер, мультимедиялық құрал, электрондық оқулықтар мен оқу-әдістемелік кешен, инновациялық ақпараттық банк, инновациялық сайт* және тағы басқалары жатады. Сондықтан да қазіргі таңда көптеген интерактивтік бағдарламалық бөлімдер, қозғалмалы объектілер құруға мүмкіндік беретін векторлық және графикалық жабдықтар көптеп шығып, қолданыс табуда.

Қазіргі заманға сай қоғамды ақпараттандыруда оқытушылардың біліктілігі, ақпараттық-коммуникациялық технологияны қолдануы – негізгі міндеттердің бірі. Осы тұрғыдан келгенде бүгінгі таңда ақпараттық технология кез келген мамандандырылған қызмет аясының, соның ішінде лингвистиканың да бөлінбес бөлшегіне айналды. Бұған дейін лингвистикалық зерттеулерде, аудармада және тіл үйретуде компьютерлер мен оларға сәйкес бағдарламаларды қолдану аса қажет болмағанымен, қазіргі уақытта студент кездерінен бастап, болашақ шетел тілдері оқытушылары, аудармашылар және лингвист-зерттеушілер өздерінің мамандандырылған қызмет аясында ақпараттық технологияларды қолдану құзырлылықтарын арттыруда.

Лингвистикадағы ақпараттық технологиялардың мүмкіндіктері филология, әдебиеттану және әлем тілдері факультетінде лингвистика, шетел филологиясы, қазақ филологиясы мамандықтары бойынша мамандандырылған студенттерге арналады. Аталмыш оқу құралы «Лингвистикалық зерттеулердегі компьютерлік технологиялар» пәнінің жұмыс бағдарламасына сәйкес жаса-

лынды. Сонымен қатар оқу құралын бакалаврларда оқытылатын «Компьютерлік лингвистика», «Лингвистикадағы ақпараттық технология» және магистратура бөлімінде оқытылатын «Тіл және жоғары технология», «Лингвистикалық зерттеулердегі жаңа технология», «Компьютерлік технология және лингвистикадағы әдіснама» сынды пәндерде де негізгі әдебиеттердің қатарына қосып, пайдалануға әбден болады.

Аталмыш еңбектің өзіндік ерекшелігі студент «Лингвистикалық зерттеулердегі компьютерлік технологиялар» пәні бойынша кредиттік жүйеге негізделіп жасалған 15 дәріс сабақтары тақырыптарына байланысты толық ақпарат ала алады. Сондай-ақ студенттің қажетті құзырлылықтарын қалыптастыруды жеңілдететін семинар тапсырмалары мен тақырыпқа қатысты пайдаланатын әдебиеттер тізімінің болуымен айқындалады.

«Лингвистикалық зерттеулердегі компьютерлік технологиялар» атты оқу құралы *«Тіл және ақпараттық технологияның негізгі ұғымдары»*, *«Ақпараттық технологияның лингвистикадағы қолданылу аясы»*, *«Компьютерлік лингвистиканың қолданбалы салалары»* секілді үш модульден және әрбір модульге қатысты дәріс тақырыптарымен қоса, әдебиеттер тізімінен тұрады. Әрбір модульде бірнеше теориялық мәселелерге негізделген тақырыптардың ғылыми тұрғыдағы түсініктемесі, семинар сабақтарында талқыланатын тапсырмалар, студенттің өз бетімен дайындалуға арналған әдебиеттер тізімі және аралық бақылау бойынша тест тапсырмалары қамтылды.

Оқу құралының соңында пән бойынша пайдаланылған әдебиеттер, интернет-ресурстар тізімі көрсетілді. Оқу құралының мазмұны мен оқу-әдістемелік аппаратында лингвистикадағы ақпараттық технологияларды толық қамтудағы базалық мүмкіндіктерді қарастыру жіті назарға алынды.

## ТІЛ ЖӘНЕ АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАРДЫҢ НЕГІЗГІ ҰҒЫМДАРЫ

### 1-дәріс. Тіл және тіл білімі

**Қарастырылатын мәселелер:** *лингвистика – тілдің қалыптасуы мен даму заңдылықтары жайлы ғылым; теориялық және қолданбалы лингвистика ұғымы; қолданбалы және компьютерлік лингвистиканың арақатынасы; тіл – таңбалар жүйесі; табиғи және жасанды тіл; жасанды тілдердің түрлері.*

**Тірек сөздер мен сөз тіркестері:** *тіл, тіл табиғаты, адамзат тілі, жан-жануарлардың тілі, ұлттық халықтық тіл, лингвистика, жеке тіл білімі, жалпы тіл білімі, тіл элементтері, фундаменталды тіл білімі, қолданбалы тіл білімі, нормативті тіл білімі, сөйлеу коммуникациясы, ақпараттық технологиялар, компьютерлік лингвистика.*

Тіл – өте күрделі, көп қырлы, сан салалы құбылыс. Ең алдымен, ол – белгілі бір адамдар қауымында қарым-қатынас жасауды жүзеге асыру үшін табиғи жолмен пайда болған өзіндік ерекше көне таңбалар жүйесі. Сонымен бірге тіл – адамдардың ойлау қабілетін іске асырушы, оны одан әрі дамытушы, ерік-жігерін, сезімін білдіруші, мәдени-тарихи дәстүрлерін ұрпақтан-ұрпаққа жеткізуші құрал.

Әлеуметтік лингвистика терминдерінің сөздігінде: тіл (ағыл. *language*) – ойлаумен, халықтың менталитетімен, мәдениетімен үздіксіз, тығыз байланыстағы адам қарым-қатынасының әлеуметтік аса маңызды құралы; шартты, бірақ құрылымдық-негізделген мағынасы бар таңбалар мен нышандардың жүйесі. Тіл – адамдар арасындағы саналы әлеуметтік сөйлеу қарым-қатынасының негізі; адамдардың мінез-құлқын реттеудің, қоғамды және оның әралуан құрылымдық элементтерін басқарудың маңызды құралы деген түсініктеме берілген [1, 241-б.].

Тілші ғалым К. Аханов: «Тіл қоғамдық құбылыстардың қатарынан орын алады. Ол – қыр-сыры мол, күрделі құбылыс. Тілдің табиғатын, мәнін танып білу үшін оның ең басты қасиетін айқындап алудың маңызы айрықша. Тілдің ең басты қасиеті делініп, әдетте оның қатынас құралы болу қызметі аталады. Тіл – адам баласы қоғамында қатынас құралы, сөйлесіп, пікір алысудың құралы ретінде пайда болды. Тарихта мұндай коллектив, алғашында, ру, тайпа, түрінде, кейінде халық тілі, ұлт болып өмір сүрді, осыған орай ру тілі, тайпа тілі, халық тілі болды. Адамдар коллективіне – мейлі ол ру немесе тайпа болсын, халық немесе ұлт болсын – осылардың бәріне де тіл қатынас құралы ретінде қызмет атқаруы оның табиғатын, мәнін айқындайтын басты қасиеті деуіміз де осыдан» – дейді [2, 38-б.].

Тілдің нақты бір «ұлттық, халықтық тіл» және «жалпы адамзат тілі» деген ұғымдары бар. «Адамзат тілі» ұғымына «жан-жануарлардың тілі» деген ұғымды қарама-қарсы қоюға болады. Тілдің, сондай-ақ ым тілі, машина тілі, өнер тілі дегендей, т.б. ұғым-түсініктері және олардың әрқайсысының өз ішіне қамтитын өзіндік мазмұн-мағыналары болады. Тілдің материалдық және мағыналық жақтары болады. Тіл өзінің даму барысында ру тілі, тайпа тілі, халық тілі, ұлт тілі, ұлттық әдеби тіл дәуірлерін басынан кешіреді. «Тіл» арқылы адамдар өмірінің әр саласындағы қолданыс қызметі де түсініледі. Мысалы, жазба тіл, ауызекі сөйлеу тілі, газет тілі, көркем әдебиет тілі, т.б. Бұл тұрғыда тіл сөз және стиль ұғымдарын беріп, солармен бір мәнде қолданылады. Алғаш рет тіл (язык) мен сөздің (речь) айырмашылықтары мен ерекшеліктерін, олардың байланысын швейцар ғалымы Ф. де Соссюр атап көрсеткен.

Тіл білімі – әлем тілдерінің құрылысын, әлеуметтік функциясын және тарихи дамуын, адамзаттың табиғи тілін, оның жалпы

зандылықтарын, қолданылуын және адам қоғамын зерттейтін гуманитарлық ғылым. Тіл білімін жеке және жалпы тіл білімі деп бөлу, зерттеудің мақсаты мен нәтижесінің негізінен туындайды. Егер бір нақты тілдің (мысалы, қазақ, орыс) немесе туыстас тілдердің (мысалы, алтай, түркі, үндіеуропа) фактілерін зерттеу мақсатын қойса, бұл жеке тіл білімінің үлесіне тиеді, ал жалпы адам баласы тілінің мәні мен құрылысы, тіл элементтерінің өзара қатысы, тілдің даму себептері мен бағыттары және тіл білімінің жалпы тілдік мәселелері қамтылса, бұл жалпы тіл білімінің зерттеу нысаны болып табылады. Тіл білімін жалпы және жеке деп бөліп қарастыру лингвистикалық зерттеулерде тілдің қоғамдық практикасына, қатысына қарай саралауға мүмкіндік береді, олар: *фундаменталды, нормативті және қолданбалы тіл білімі*.

*Фундаменталды тіл білімі* – тілдің қазіргі күйін, жағдайын, бұрынғы өткен тарихын зерттейді. Негізгі мақсаты – тілге ғылыми талдау жасау емес, сол күйді бақылап, тәжірибе жүзінде жинақтау, тілдің (бүгінгі, кешегі) күйі қандай деген сұраққа жауап беру.

*Нормативті тіл білімі* – фундаменталды тіл білімі жинақтаған зерттеулер негізінде жасалынады, тілдің қоғамдық қызметі, стиль тармақтарының қолданылуы, тілдік нормалардың қалыптасуы мен өзгеруі, тілдің біркелкі қалыптасқан нормаларын қолдану шарты және т.б. тіл саясатының жүзеге асырылуы тіл нормаларының қалыптасуынан көрінеді.

*Қолданбалы тіл білімі* – тіл нормалары толық қалыптасқан жағдайда ғана жүзеге асады, ол тілдік нормаларды таратып қана қоймайды, норманы жетілдіріп, тұрақтандырады. Қолданбалы тіл білімінің негізгі мақсаты – сөйлеудің жаңа қолданыстарын енгізу, сөйлеу коммуникациясының жаңа техникалық жағын ойластыру, тіл нормаларын үйрету, қарым-қатынас құралдарының жаңа түрлерін тарату және сөйлеу коммуникациясы қалай жасалады деген сұраққа жауап беру.

Тіл білімінің бір-бірімен тығыз байланысты негізгі үш саласы бар:

- фонетика – тілдің дыбыстық жағын;
- лексикология – тілдің сөздік қорын;
- грамматика – тілдің морфологиялық және синтаксистік құрылымын зерттейді.

Тілдің қалпын дамудың белгілі бір мезгілінде өзара үйлесімді және байланысты элементтер жүйесі ретінде қарастыру синхронды тіл білімінде, тілдің тарихи дамуын, жеке тіл деректерін және тіл жүйесін бүтіндей зерттеу саласы диахронды тіл білімінде зерттеледі. Сипаттамалы тіл білімі – мәдени дамуы бір-біріне ортақ, бір-біріне ұқсамайтын тілдік жүйе құрылымын, ал салыстырмалы-тарихи тіл білімі фонетикалық сәйкестіктерді, морфологиялық ұқсастықтар мен варианттар, туыстас тілдер этимологиясы, классификациясы, реконструкциясы және т.б. мәселелерді бір тілдік семьяға жататын тілдер негізінде зерттейді. Түркі тілдерінің салыстырмалы-тарихи тіл білімі XI ғасырда өмір сүрген ғұлама энциклопедист ғалым М. Қашқаридың «Диуани лұғат ат-түрк» еңбегінен басталады.

Тіл білімі мен басқа ғылымдардың тоғысуы нәтижесінде лингвистиканың жаңа салалары дүниеге келді. Тіл білімі мен әдебиеттану ғылымы стилистика ғылымының негізін салды. Тіл білімі мен әлеуметтану – *әлеуметтік тілі білімі*, тіл білімі мен психология – *психоллингвистика*, тіл білімі мен этнография – *этнолингвистика*, тіл білімі мен мәдениеттану – *лингвистикалық мәдениеттану* (лингвомәдениеттану), сонымен қатар *нейролингвистика*, *лингвосемиотика*, *когнитивті лингвистика*, *қолданбалы лингвистика*, *компьютерлік лингвистика* және т.б. интеграциялық және дифференциалдық үдерістердің нәтижесінде пайда болған тіл білімінің жаңа салаларын атап өтуімізге болады.

Ғылымдар тоғысы нәтижесінде пайда болған *қолданбалы лингвистика* (ағыл. *Apply Linguistics*) машиналық аударма жасау, автоматты түрде ақпарат іздеу, жазуды ойлап табу, оны жетілдіру, ауызша сөйлеудің транскрипциялық жүйесі мен бөгде тілден кірген сөздердің транслитерациялық жүйесін құру, көзінде ақауы бар адамдарға арналған жазу жүйесін жасау, жаңа бұйымдар мен зат атауларының даму үдерістерін зерттеу, қалыптасу ережелерін үйрету және арнайы лингвистикалық анықтамаларды құру тәрізді қазіргі заманғы практикалық міндеттерді шешу үшін оның мүмкіндіктерін пайдалану негізінде тіл теориясы мәселелерін қалыптастырудағы тіл білімінің бағыты. Қолданбалы лингвистика термині XX ғасырдың 20-жылдарының соңында пайда болды.

Компьютерлік лингвистика / *Computational Linguistics* (ағыл. computer < лат. *computare санау, есептеу*) – қолданбалы лингвис-



тиканың құрамдас бір бөлігі. Арнаулы және басқа компьютерлердің әртүрлі интеллектуалды қызметтерде кең түрде қолданылуына орай дүниеге келуі және дамуы. Компьютерлік лингвистиканың негізгі зерттеу мақсаты – компьютерлендірудің лингвистикалық мәселелерін шешу, яғни ЭЕМ-мен жұмыс істеуге арналған жасанды тілдер құрылымын анықтау, ақпараттық банктерді құру, қалыптастыру. Мәтіндерді өңдеу үшін машинамен сөздіктерді, алгоритмдерді құрастыру және әртүрлі тілдік үдерістерді модельдеу және басқа мәселелерді табиғи тілдің семантикалық заңдарын ескере отырып, шешу [3, 186-б.].

Лингвистикадағы ақпараттық технологиялардың қолданылу мүмкіндіктерін оқу аталмыш ғылым саласына сәйкес негізгі ұғымдарды білуді бағамдайды. Олардың арасынан *лингвистика* (тіл, лингвистика, компьютерлік лингвистика және т.б.) мен *информатика* (ақпарат, алгоритм, модель және т.б.) саласына жататын ұғымдарды бөліп қарастыруға болады. Бұл ұғымдармен танысуды лингвистер, шет тілдерінің оқытушылары және аудармашылардың тікелей қызметтерінің кәсіби жағын сипаттайтын лингвистикалық терминдерден бастаған жөн.

Лингвистика немесе тіл білімі – дәстүрлі түрде адамзат тілін зерттейтін ғылым [4, 28]. Лингвистер тілдің қалыптасуы (одан фонетикалық, лексикалық, грамматикалық деңгейі мен мәтін деңгейін ерекшелеу), тілдің әлеуметтік түрленуі, тілдік айтылымдардың туындауы, түсіну сұрақтары, әртүрлі типті қоғамдағы тіл қызметінің принциптері, тілдің пайда болуы, дамуы және оның басқа аспектілеріне қатысты мәселелермен жұмыс істейді [5, 618 – 622-б.].

Тілдің зерттелу аспектісіне, ұлттық салт-дәстүрі мен ғылыми әдістемесіне сәйкес лингвистиканың әртүрлі салалары жеке-жеке қарастырылады, мысалы, құрылымдық лингвистика, әлеуметтік лингвистика, психоллингвистика және т.б.

Ақпараттық технологияларды қолданумен тығыз байланысты лингвистиканың бөлімін анықтау үшін теориялық және қолданбалы лингвистиканың арасындағы айырмашылықтарды білген жөн.

Теориялық немесе фундаменталды лингвистика – бұл жеке тілдік жағдаятты объективті анықтауға, оның тарихы мен заңдылықтарына негізделген лингвистика саласы. Лингвистиканың бұл саласы «Тіл қандай?» деген сұраққа толық жауап береді [6, 214 – 215-б.].

Қолданбалы лингвистика XX ғасырдың 20-жылдарының соңынан бастап дамып келеді. Тіл білімінің бұл саласы тілді қолдану арқылы практикалық тапсырмаларды шешуге арналған әдістерді құрастырумен айналысады [13, 397-б.]. Қолданбалы лингвистика «Тілді қалай жақсы/дұрыс қолдануға болады?» деген сұраққа жауап береді.

Ресейде және шет елдерде қолданбалы лингвистика ұғымы жайлы түрлі интерпретациялар қалыптасты. Егер шет елдерде 1930-1940 жылдары қолданбалы лингвистика (Applied Linguistics) ретінде, ең алдымен, шет тілін оқыту үдерісі, оны оқыту әдістемесі, оқыту мақсатындағы грамматиканы сипаттау ерекшеліктері ұғынылса, Ресейде 1950 жылдардан бастап қолданбалы лингвистика компьютерлік технологиялармен және ақпаратты өңдеудің автоматтық жүйелерімен ұқсас қарастырылып келеді [7, 6-б.]. Осыған байланысты орыс ғылымында қолданбалы лингвистика компьютерлік / есептеуіш / автоматты / инженерлік лингвистикаға синоним түрінде қаралады.

Ғылым дамуының қазіргі кезеңінде қолданбалы лингвистика аясында тілді қолдануды оңтайландыру бойынша бірнеше бағыттар айқындалды. Зерттеушілер оларды дәстүрлі («мәңгілік») және жаңа деп екі үлкен топқа біріктірді.

Қолданбалы лингвистиканың дәстүрлі бағыттары мен міндеттері:

- жазуды құру және жетілдіру;
- ауызша тілдің транскрипциялық жүйелерін құру;
- шет тілдік сөздердің транслитерациялық жүйелерін құру;
- стенография жүйелерін құру;
- загіптарға арналған жазу жүйелерін құру;
- ғылыми-техникалық терминологияны бірыңғайлау, біріздендіру және стандарттау;
- жаңа бұйымдар, тауарлар, химиялық қосындылардың атауларының жасалым ережелерін құру және үдерістерді зерттеу;
- мәтіндерді шет тілдік формаға (аударма) дұрыс аудару әдістемесін құру;
- тілдерді оқыту әдістемесін жетілдіру және т.б. [5, 397-б.].

Қолданбалы лингвистиканың жаңа міндеттері:

- машиналық аударманың лингвистикалық негіздерін жасау;
- құжаттарды автоматты индекстеу және аннотациялау;

- *мәтіннің автоматты анализі;*
- *мәтіннің автоматты синтезі;*
- *ақпаратты автоматты іздеуге арналған сөздік-тезаурустарды жасау [8].*

1960 жылдары ерекше ғылыми бағыт ретінде қалыптасқан компьютерлік лингвистика пәні тілді қолданудың практикалық міндеттерін компьютерлер көмегімен шешуге байланысты қолданбалы лингвистиканың салаларын қарастырады.

Компьютерлік лингвистика – компьютерлік құралдардағы бағдарламаларды, кез келген жағдайдағы тіл қолданысын модельдеуге арналған мәліметтерді ұйымдастыру мен өңдеу технологияларын, сондай-ақ лингвистика және онымен шектес пәндерде тілдің компьютерлік моделін қолдану саласы [7, 13-б.].

Компьютерлік лингвистикада тілдің күрделі құрылымына сәйкес тілдің жеке деңгейлерімен қатар қойылатын, тілдік хабарламалардың туындауы мен қабылдану үдерістері немесе тілге байланысты басқа да адамзаттық қызмет түрлеріне қатысты түрлі бағыттар қалыптасып, дамуда. Компьютерлік лингвистиканың бағыттары:

- *мәтіннің автоматты анализі;*
- *мәтіннің автоматты синтезі;*
- *автоматты сөздіктерді құру және қолдау;*
- *автоматтандырылған ақпараттық-іздеу жүйелерін құру;*
- *машиналық аударма;*
- *тілді оқытудағы автоматты жүйелерді құру;*
- *анонимді мәтіндерді автоматты атрибуциялау мен дешифрлеу;*
- *лингвистикалық деректер қорын құру;*
- *теориялық және қолданбалы лингвистиканың мәселелерін шешуге байланысты бағдарламалық құралдарды дайындау және т.б. [9; 53-б.].*

Жалпы лингвистика, сондай-ақ жекелей алғанда компьютерлік лингвистика да әртүрлі типтегі тілдермен, олардың жеке деңгейлерімен жұмыс істейді. Тіл жалпы түрде кейбір социумда қарым-қатынас үшін қолданылатын таңбалар жүйесі ретінде анықталады [5, 604-б.; 29, 5-б.].

Тіл табиғи және жасанды түрге бөлінеді. Табиғи тіл – тарихи қалыптасқан, белгілі этникалық топ немесе ұлттық мемлекет-

те қолданылатын таңбалар жүйесі. Бұл тіл үндіеуропа тілдік семьясына жататын орыс, ағылшын тілдері немесе фин-угор тілдік семьясына жататын фин, эстон тілдерін қамтиды.

Жасанды тілдер табиғи тілді қолдану тиімсіз немесе мүмкін емес салаларда жасанды түзілетін таңбалар жүйесін білдіреді. Жасанды тілдер өз ішінде мамандандырылмаған немесе халықаралық және мамандандырылған *эсперанто*, *волапюк* және т.б. тілдер болып бөлінеді. Соңғысына XVI ғ. қалыптаса бастаған математика, логика, химия және т.б. ғылым тілдері және 1940 жылдардан бастап адам мен компьютер арасындағы диалогті жеңілдетуге байланысты адамзат қызметінің арнайы салаларында таралған адам-машиналық қарым-қатынас тілдері жатады [5, 201-202-б.].

Адам-машиналық қарым-қатынас тілдеріне мысал ретінде қарапайым символикалық кодтау жүйесінен бастап (ассемблерлер) мамандандырылған бағдарламалау тілдері *C++*, *Java*, *Python*, *ErLang* және т.б. алынады. 1980 жылдарға қарай әлемде 500-ге жуық бағдарламалау тілдері бар болған [5, 202]. Бағдарламалау тілдерінің жалпы саны бірнеше мыңға жетсе де, қазіргі кезде бастапқы мөлшері белсенді қолданыста жүр [10]. Бұл фактілер адам-машиналық қарым-қатынас мәселелерінің өзектілігі мен оны шешудің көптеген жолдарының бар екендігін көрсетіп отыр.

Демек, лингвистика – табиғи адамзаттық тілдің пайда болуы, қалыптасуы мен қызметінің заңдылықтары жайлы ғылым. Лингвистика мен компьютерлік лингвистиканың пәні оның ерекше бөлімі түрлі мақсаттарда қолданылатын тіл таңбалар жүйесі болып табылады.

### Семинар тапсырмалары

1. Тіл дегеніміз не? Тілге қатысты қандай анықтамаларды білесіз және оларды салыстырып көріңіз. Анықтамалардағы тірек сөздерді бөліп қараңыз. Қайталанатын тірек сөздерді негізге ала отырып, өз анықтамаңызды құраңыз.
2. Тіл білімі дегеніміз не? Оның салаларын атап, логикалық кесте түрінде беріңіз.
3. Лингвистиканың қай саласының ақпараттық технологияларға қатысы бар? Дәлелдеңіз.
4. Қолданбалы және компьютерлік лингвистиканы синоним ретінде қарастыруға болады ма? Өз жауабыңызды дәлелдеңіз.
5. Компьютерлік лингвистиканың негізгі бағыттарын атаңыз. Бір бағытына арнайы тоқталып өтіңіз.

6. Тілдің түрлі анықтамаларын салыстырыңыз. Тілді толық көрсететін анықтаманы шығарып беріңіз.
7. Сіз табиғи, жасанды тілдердің қандай түрлерін білесіз? Табиғи және жасанды тілдердің әр түріне мысал келтіріңіз. Компьютерлік лингвистика қай тілде жұмыс жасайды?

### Әдебиеттер:

#### *Негізгі:*

1. Аханов К. Тіл білімінің негіздері. – Алматы, 2004. – 38 б.
2. Березин Ф.М., Головин Б.Н. Общее языкознание. – М.: Просвещение, 1979.
3. Рождественский Ю.В. Лекции по общему языкознанию. – М.: Высшая школа, 1990.
4. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. – 3-е изд. – М.: ЛКИ, 2007.
5. Прикладное языкознание: учебник / отв. ред. А.С. Герд. – СПб.: СПбГУ, 1996.
6. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
7. Kinnerlesley B. The Language List. Collected Information On About 2500 Computer Languages, Past and Present. <http://people.ku.edu/~nkinneres/LangList/Extras/langlist.htm> (қаралған күн: 28.02.2012).
8. Лингвистический энциклопедический словарь / Гл. ред. В.Н. Ярцева. – М.: Сов. энциклопедия, 1990. – 685 с.
9. Тіл білімі сөздігі. Словарь по языкознанию / құраст. Э.Д. Сүлейменова, Г.Б. Мәдиева, Н.Ж. Шәймерденова, Р.Ә. Авакова, т.б. – Алматы, 1998.
10. Сүлейменова Э.Д., Шәймерденова Н.Ж., Смағұлова Ж.С., Ақанова Д.Х. Әлеуметтік лингвистика терминдерінің сөздігі. – Алматы, 2008. – 241 б.
11. Большой энциклопедический словарь (БЭС). Языкознание. – М.: Большая Российская энциклопедия, 1998.

#### *Қосымша:*

1. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. – СПб.: Книжный Дом, 2007. – С. 36-40.
2. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд., испр. – М.: Флинта: наука, 2007. – С. 63-64.

## **2-дәріс. Ақпарат. Ақпараттық технологиялар**

**Қарастырылатын мәселелер:** *ақпарат – информатика мен кибернетиканың зерттеу пәні; лингвистикадағы ақпараттық технологиялар ұғымы; ақпарат түрлері; ақпаратты кодтау мен оның жеткізушілері; ақпараттық төңкерістер; информатикадағы модель мен алгоритм ұғымы; жасанды интеллект ұғымы.*

**Тірек сөздер мен сөз тіркестері:** *технология, ақпараттық технология, ақпарат, ақпарат түрлері, ақпараттық үдеріс, жаңа ақпараттық технология, компьютерлік ақпараттық технология, ақпараттық үдеріс, ақпараттық операциялар, ақпараттық төңкеріс, сандық ақпарат, жасанды интеллект.*

*Технология* – материалды өңдеу үдерістерінің жиынтығы немесе материалды өңдеу тәсілдері туралы білім жиынтығы. Технология – түрлендіруді жүзеге асыруға мүмкіндік беретін тәсіл. Бұл бізге технология түсінігін тек материалды өндіріс аясында ғана емес, сонымен қоса адам іс-әрекетінің басқа да аясында да, соның ішінде білім беру жүйесінде де пайдалануға мүмкіндік береді. Технология ұғымының 300-ге жуық анықтамасы бар. Ақпараттық технологияның арқасында әрбір адамның тек дәстүрлі ақпарат көздерімен ғана емес, сонымен қатар дәстүрлі емес ақпарат көздерімен жұмыс істеуге мүмкіндіктері бар.

*Ақпарат* (лат. *informatio* баяндау, мәлімет) – айналадағы дүние туралы анықталмағанды (белгісізді) анықтайтын мәліметтер. Олар сақтау, түрлендіру, беру және пайдалану объектісі болып табылады.

Кибернетика ғылымының атасы Ноберт Винер: «Ақпарат – сыртқы дүниеден біздің оған бейімделуіміздің және біздің сезімдеріміздің бейімделуінің үдерісінде алынған мазмұнды көрсету», – дейді.

*Ақпараттық технология* – ақпаратты өндеумен байланысты үдерістер. Ақпараттық технология – ақпаратты өндеу үдерісі арқылы жүзеге асатын құралдар мен әдістер жиынтығы.

Ақпарат түрлерін төмендегідей топтастыруға болады:

***Ақпараттың берілуіне қарай:***

- таңбалы – мәтіндік (әріп, цифр, таңбалар, т.б.);
- графикалық (суреттер, бейнелер көмегімен, т.б.);
- дыбыстық (тіл лексемалары ауызша немесе жазба түрінде аудиалды жолмен берілетін);
- видеоақпарат (видеожазба түрінде берілетін).

***Пайда болу және қолдану аумақтарына байланысты түрлері:***

- биологиялық;
- әлеуметтік;
- ғылыми;
- экономикалық, т.с.с.

***Берілу және қабылдау тәсілдеріне байланысты түрлері:***

- визуалды (таңбалар мен бейнелер арқылы);
- аудиалды (дыбыс арқылы);
- тактильді (сезім арқылы);
- дәм сезу;
- иіс сезу;
- машиналық (есептеу техникасының құралдары арқылы).



**1-сурет.** Ақпараттық технология мүмкіндіктері

Кез келген ұйымда басқару қызметтері нәтижелі ақпараттың, қорытылған мәліметтердің қайта құру технологиясына негізделеді. Технология ұғымының өзі өнеркәсіптік өндірісте қолданылады, өндіріс үдерісінде өнімді әзірлеу тәсілдерінен және өзара байланысқан материалдарды өңдеу жүйесі ретінде анықталады. Әдістер жүйесі техникалық құралдарды пайдалану негізінде ақпараттық жинау, жіберу, өңдеу, жинақтау, сақтау тәсілдерін қолдану арқылы ақпараттық технология анықталады.

Ақпараттық технологиялар ақпараттық үдерістердің әр қилылығына байланысты мынандай технологияға жіктеледі:

- ақпаратты жинау;
- ақпаратты жеткізу;
- ақпаратты өңдеу;
- ақпаратты жинақтау;
- ақпаратты сақтау;
- ақпаратты ұсыну;
- ақпаратты пайдалану;
- ақпаратты қорғау.

Нақты ақпараттық технологияны жүзеге асыру үшін мыналар орындалуы қажет:

- ақпараттық үдерістің өзін жүзеге асыратын тиісті техникалық құралдардың кешені;
- техникалық кешенмен басқару құралдарының жүйесі;
- техникалық құралдардың барлық іс-әрекеттерінің жүзеге асырылуын байланыстыратын ұйымдастырушылықты әдістемелік қамтамасыз ету.



2-сурет. Ақпараттық технология сұлбасы

Қазіргі заманда ақпараттық үдерістердің әр түрлерін жинақты түрде пайдалануға болады. Жаңа ақпараттық технологияның дамуы компьютерлік ақпараттық технологиямен тығыз байланыста екендігінде. Қоғамның қазіргі заман дәуіріндегі ғылыми-техникалық прогрестің дамуына байланысты адам іс-әрекетінің барлық аясында компьютерлік және желілік технологиялардың бірігуінен пайда болған ақпараттық технологияны пайдалану кең етек алуда. Қоғамды компьютерлендіру, техника құралдарының белсенді түрде енуі бірқатар көкейкесті міндеттерді шешуде. Осыған орай қазіргі таңда *жаңа ақпараттық технологиялар*, ЭЕМ-дер мен олардың желілері арқылы, әсіресе дербес компьютерлер көмегі арқылы ақпаратты дайындау, жинау, жеткізу және өңдеу технологиялары жүзеге асуда.



**Ақпараттық үдерістер** – адамдар арасында, тірі ағзаларда, техникалық құрылғыларда және қоғамдық өмірде ақпаратты жеткізу, жинақтау, түрлендіру үдерістері. **Жаңа ақпараттық технологиялар** – ЭЕМ жадында сақталған картотекадағы, каталогтардағы, әртүрлі мұрағаттар мен кітапханалардағы ақпараттарды жинақтауға, біріктіріп сақтауға, керектілерін жылдам іздеп табуға болатын әртүрлі мәліметтер базасы мен ақпараттық жүйелер. Жаңа ақпараттық технологиялардың дамуын ЭЕМ желілеріне негізделген электрондық поштасыз, байланыс желілері мен ақпараттық коммуникацияларсыз көзге елестету мүмкін емес.

Тілдің негізгі міндеттерінің бірі – оның адамдар арасындағы ақпарат жеткізу құралы ретінде қолданылуы. Сондықтан да тіл жайлы айтқанда ақпарат ұғымын айналып өтуге болмайды.

Ақпарат күнделікті өмірде қоршаған ортадағы істердің жағдайлары, оның құрамы мен ондағы өтіп жатқан үдерістер туралы мәлімет деп түсініледі [1]. Ақпаратты оқытатын арнайы ғылымдарда бұл ұғым өзгеше анықталады: кейбір хабарламаларды мағыналық мазмұнсыз (ақпаратты жеткізу теориясы) кодтайтын кейбір алфавиттердің сигналдары мен символдарының тізбегі немесе ішкі әлемнен алынған және қоршаған ортада (кибернетикада) тірі ағзаға (немесе техникалық жүйеге) адекватты түрде әсер етуге мүмкіндік беретін мазмұн [2, 11 – 12-б.].

Ақпараттың түрлі анықтамаларын жинақтай келе, бұл ұғымның келесі анықтамасын ұсынуға болады: **ақпарат** – қоршаған ортадағы жүріс-тұрысын реттеу үшін адам, тірі ағзалар немесе техникалық жүйелер арқылы жеткізілетін қоршаған әлем туралы мәліметтер.

Қазіргі қоғамдағы ақпараттың рөлі өте зор. Тіл арқылы кодталатын ақпарат ілімге айналады; ал ілім ұрпақтан-ұрпаққа жеткізіледі, сол арқылы қоғамдық ұстанымдардың сабақтастығын қамтамасыз етеді.

Ақпарат вербалды немесе бейвербалды болып кодтала алады. Ақпараттың кодталу тәсілдерінің әртүрлілігі оны ұсынудың көптеген тәсілдерін көрсетеді:

- мәтіндер;
- суреттер, сызбалар, фотосуреттер;
- сәулелік немесе дыбыстық сигналдар;
- электрлік немесе жүйкелік импульстер;

– *ым-ишара*;  
 – *иіс пен дәм сезу*;  
 – *ағзаның белгілері мен қасиеттері ұрпаққа берілетін хромосомалар және т.б.*

Мысалдар көрсеткендей, ақпаратты ұсынудың тәсілдері көп. Алайда адам ақпаратты жеке сезім мүшелері көмегімен ғана қабылдай алатындықтан, осы негізде ғана ақпарат түрлерін мақсатты түрде жіктеу керек. Қоршаған әлем жайлы ақпараттың қандай сезім мүшелерімен қабылданғандығы және қандай сигналдар жүйесімен кодталғанына қарай дыбыстық, дәмдік, тактильді, визуалды-образдық және визуалды-символикалық ақпаратты ажыратуға болады. Әсіресе ақпараттың соңғы екі түрі қазіргі адам үшін мейлінше маңызды. Егер XX ғасырда адамдар визуалды-образды түрімен жұмыс істесе, XXI ғасырда визуалды-символикалық ақпарат маңыздырақ бола бастады.

**Символ** (грек. *symbolon*) – кейбір зат немесе құбылысты білдіретін таңба. Лингвистикада *символ* деп қоршаған әлемнің заттары немесе құбылыстарын білдіре алатын минималды бірлік, яғни сөзді айтады. Информатикада әріптер, тыныс белгілері, сандар мен басқа да баспа мәтіндерінің таңбалары, сонымен қатар түрлі табиғи, жасанды тілдердің әліпшесі мен фонетикалық жүйесін құрайтын ауызша мәтіннің дыбыстық таңбалары – фонемалар символ деп есептелінеді. Бұл символдар жеткізілетін ақпаратты кодтайтын, сөздер мен сөйлемдерге айналады.

Ақпаратқа қатысты белгілі бір амалдармен байланысты операциялар **ақпараттық операциялар** деп аталады. Қазіргі таңда ақпараттарға келесі амалдарды жүзеге асыруға болады:

құру	қабылдау	комбинациялау
сақтау	жеткізу	көшіру
іздеу	қабылдау	формалдау
өлшеу	қолдану	бөлшектерге ажырату
оңтайландыру	бұзу	өңдеу
жинақтау	тарату	қайта жасау

Қоғам дамуының белгілі деңгейінде адамдар қолданатын ақпарат көлемінің әрдайым көбеюіне байланысты оны өңдеу мен сақтау үшін арнайы техникалық құралдарды тарту қажет болды.

Жаңа техникалық құралдарды ойлап табумен байланысты ақпаратты фиксациялау мен жеткізу тәсілдеріндегі принципті өзгерістер **ақпараттық төңкеріс** деген атау алды. Зерттеушілер үш ақпараттық төңкерісті бөліп қарастырады:

1. Шамамен б.з.д. 3000 жыл бұрын жазудың пайда болуы (шумер жазуы): ақпарат жинақталуы мүмкін.

2. 1453 ж. – кітап басудың пайда болуы: ақпарат көпшілікке қолжетімді бола бастады.

3. 1960 жылдардың басы – дербес ЭЕМ мен телекоммуникациялық желілердің пайда болуы: ақпарат жоғары жылдамдықта автоматты өңделіп, электронды түрде жеткізіле алатын болды.

Үшінші ақпараттық төңкерістің ерекшелігі: XX ғасырдың ортасында ақпаратты зерттейтін арнайы ғылымдар: информатика мен кибернетиканың пайда болуы. *Информатика* – ЭЕМ көмегімен ақпараттың жинақталуы, өңделуі мен жеткізілуі жайлы ғылым. Ақпаратты басқару, байланыстыру мен қайта өңдеу жайлы ғылым кибернетика деп аталады.

Техникалық құралдар көмегімен жүзеге асатын, ақпараттық алмасуды кескіндейтін ақпарат теориясы (байланыстың математикалық теориясы) аясында К. Шеннон мен У. Уивер модель ұсынды (3-сурет).



3-сурет. К. Шеннон мен У. Уивер моделі [3, 131-6.]

Ақпараттық технологиялар үшін берілген модельдегі кодтау мен декодтау құрылғысының нұсқауы ерекше маңызға ие. Себебі информатиканың маңызды міндеттерінің бірі «адамзаттық» символдарда кодталған ақпаратты компьютерге түсінікті ақпаратқа және керісінше ақпарат алмастыру болып табылады.

Ақпараттық алмасуда компьютерлер символдық ақпараттың үлкен массивтерін кодтау, өңдеу, сақтау мен жеткізу құралына айналады. Компьютер көмегімен ақпаратты алу, сақтау, жеткізу, тарату мен қайта жасау заңдылығы, әдістері мен құралдарының жинақталуы «*ақпараттық технологиялар*» атауына ие болады.

Бұл ұғымның кәсіби саладағы (лингвистика) қолданысы үшін «*лингвистикадағы ақпараттық технологиялар*» тіркесімін аламыз. Яғни компьютер көмегімен тіл және оның қызметінің заңдылықтары жайлы ақпаратты алу, сақтау, жеткізу, тарату, қайта жасау заңдылықтарының, тәсілдері мен құралдарының жиынтығы [4, 8-б.].

Бұл ғылым саласы міндеттерінің бірі адам мен компьютер арқылы ақпараттың кодтау тәсілдерін салыстыру болып табылады.

Ақпаратты белгілі бір алфавит арқылы ұсынуды *кодтау* деп атайды. Бір белгі тобынан екінші белгі тобына көшіру ережесін *код* деп атайды. Ақпаратты сақтау, қабылдау, ұсыну және өңдеу әдістері іс жүзінде ақпараттың ұсынылу (кодталу) түріне байланысты болады. Ақпараттың кодталуы кейде шифрлау деп те айтылады.

*Кодтау* дегенде шартты түсініктемелердің тізбегі түріндегі ақпараттың ұсынылу үдерісі түсініледі. Басқа сөзбен айтқанда, кодтау – объектілер және олардың арасындағы қарым-қатынастар мен белгілі бір тілдің символдары мен сөздерінің қатар қойылуы [2, 39 – 40-б.].

Кодтау үдерісінде сөздер (символ) мен оның мағынасының өзара қарым-қатынасы, әдетте *семантика* деп, ал тілдің дұрыс сөйлемдерін құру немесе сипаттауға мүмкіндік беретін сөздер мен сөздер тобының ортақ синтаксистік қасиеттерін білдіретін ережелер *грамматика* деп аталады [5, 98-б.; 51, 19-б.].

Адамның ақпаратты кодтауының тәсілдері жоғарыда айтылды. Компьютер ақпараттың барлық танымал түрлерін өңдей алады:

- *сандық;*
- *әріптік (вербалды);*
- *графикалық;*
- *дыбыстық;*
- *бейнеақпаратты.*

Компьютердегі ақпарат әліпбиі екі саннан тұратын (0 мен 1) екілік кодта ұсынылады.

Осылайша, адам пайдаланатын *сандық ақпараттағы* ондық санау жүйесі ЭЕМ-да келесі 0 мен 1 символдарының тіркесімен келеді:

0 – 0	4 – 100	8 – 1000
1 – 1	5 – 101	9 – 1001
2 – 10	6 – 110	10 – 1010
3 – 11	7 – 111	

Компьютермен *вербалды ақпаратты* кодтау үшін бастапқыда ASCII коды (American Standard Code for Information Interchange) қолданылды. Бұл кодта бір символды кодтауға 1 байт (немесе 8 бит) қажет етіледі. ASCII-де 256 символды кодтауға болады. Әрбір символға сәйкес 0-ден 255-ке дейінгі ерекше ондық код қойылады. Мәселен, екілік кодта «КОМПЬЮТЕР» сөзінің жазбасы келесі түрде көрінеді (1-кесте).

*1-кесте*

**КОМПЬЮТЕР сөзін құрайтын символдардың  
екілік коды [5, 62-б.]**

1	2	3	4	5	6	7	8	9
К	О	М	П	Ь	Ю	Т	Е	Р
10001010	10001110	10001100	10001111	10011100	10001110	10010010	10000101	10010000

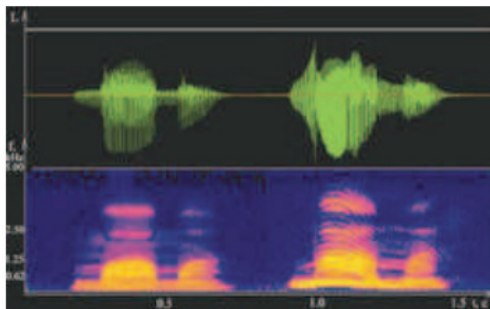
Компьютер жадында берілген сөзді кодтауға 9 сегізбитті сандардың комбинациясы, яғни, 9 байт қажет етіледі. Кіші әріппен жазылған осы сөздің екілік коды өзгеше болатынын еске ұстау керек.

Қазіргі кезде кодтаудың сол бір жүйеде шифрлана алатын символдардың көлемінің көбеюі үшін бір символды кодтауға екі байт қолданылатын UNICODE стандарты пайдаланылады.

*Графикалық ақпаратты* кодтауға, әдетте екі тәсіл қолданылады: суретті растрлық немесе векторлық көрініс түрінде ұсыну. Растрлық көрініс нүктелерден тұратын (пиксельдер) тармақтардан құралады. Векторлық көрініс – қарапайым графикалық объектілерден тұратын графикалық объект. Мысалы, қималар мен до-

ғалар. Осы объектілердің орналасуы нүктелердің координаттары мен радиустың ұзындығымен анықталады.

Дыбыстық ақпаратты кодтау бұл ақпараттың материалды сипатына арқа сүйейді. Дыбыс үздіксіз өзгертін амплитуда мен дыбысталу жиілігінен тұратын дыбыстық толқындарды білдіреді. Сигналдың амплитудасы үлкен болған сайын, ол қаттырақ, сигналдың жиілігі көп болған сайын, тоны жоғары болады. Дыбыстық толқынды визуалды елестетуге фонограмма көмектеседі. Яғни, ол арнайы құралдармен анықталады және бейнеленеді. Мысалы, монитор экранындағы дыбыс толқындарының тербелісі (2-сурет).



4-сурет. Мама сөзінің визуалды көрінісі [4]

Бейнежазбаны дыбыстық ақпаратқа кодтауда уақыт өсінде бірінен екіншісіне баяу өтетін жеке кадрлар түріндегі визуалды көріністер қосылады.

Лингвистикалық мәліметтерді компьютерлік өңдеуде компьютерлік лингвистикалық модель және лингвистикалық мәселелерді шешудің алгоритмі жайлы білу қажет.

**Модель** деп реалды заттар мен құбылыстарды алмастыратын және олардың тек шынайы белгілерін қосатын, кейбір заттар немесе құбылыстардың материалды немесе идеалды образдары есептелінеді [7, 38-б.]. Материалды модельдердің мысалы ретінде химияда молекулалардың суреттері немесе үшөлшемдік бейнелері, астрономияда күн жүйесі, анатомияда адам ағзасы алынады.

Лингвистикалық модельдер көп жағдайда тілдің құрылымы мен қызметінің ерекшеліктерін, сөз бен мәтінді құру мен қабылдауды айқындауға мүмкіндік туғызатын идеалды құрылым болып табылады [4, 14-б.]. Қарапайым лингвистикалық модельдер фонемалардан сөздің, есім мен етістікті топтардан сөйлемнің, сюжет бірліктерінен мәтіндердің құрылымын бейнелейді. Осылайша, сюжеттік грамматикада мәтіннің базалық элементтері ретінде экспозиция, оқиға мен эпизод ұсынылады [8, 27-б.]. Күрделі лингвистикалық модельдер әртүрлі деңгейлерді құрайтындардың ең үлкен санын қосады және кешендік мақсаттармен ерекшеленеді (2.4-параграфтағы *Мәтіннің автоматтық анализі және синтезі* тараушысымен салыстырыңыз).

Компьютерлік лингвистикалық модельдердің құрылуы әрекеттердегі тізбектіліктің орындалуына байланысты. Берілген мәселелерді шешуге бағытталған осындай әрекеттердің кезектерін формалды түрде сипаттау *алгоритм* деген атауға ие [7, 40-б.]. Алгоритмдер вербалдық нұсқаулықтар, блок-схемалар, кестелер немесе бағдарламалау тілдерінде жазылуы мүмкін. Түрлі сипаттағы алгоритмдердің мысалдарын еңбектен қараңыз [4, 18 – 19-б., 36 – 37-б.].

1970 жылдардан бастап адамдардың әртүрлі аядағы және пәндік саладағы адам қызметін модельдеудің түрлі жолдары жасанды интеллект құру жұмысымен бағытталып келеді. **Жасанды интеллект** (ағылш. *Artificial Intelligence*) деп күрделі адам-шиналық және роботты-техникалық жүйелерді жасауға қатысты зерттеулердің пәнаралық саласын айтуға болады [9, 14-б.].

Берілген тараушаның мазмұнына қорытынды жасай келе, айтарымыз: адамзат қоғамының өмір сүруінің бұлжымас шарты болып табылатын ақпарат қоршаған ортадағы өзгерістерге адекватты реакция беру үшін адам, тірі ағзалар немесе техникалық жүйелер арқылы жеткізілетін әлем туралы деректерді білдіреді. Ақпаратты алу, сақтау, жеткізу, тарату мен қайта жасауды жүзеге асыратын компьютерлік құралдар, сондай-ақ сәйкес заңдылықтар мен тәсілдер ақпараттық технологиялар мағынасын иеленді. Егер біз компьютер арқылы тіл мен оның қызметінің заңдылығына қатысты кез келген ақпаратты қабылдап, сақтап, жеткізіп, таратсақ, лингвистикадағы ақпараттық технологиялармен байланыста боламыз.

### Семинар тапсырмалары

1. Технология, ақпараттық технология, компьютерлік технология, жаңа технология, жоғары технология терминдеріне түсініктеме беріңіз. Сіз аталған атауларды қалай пайдаланасыз?
2. Ақпараттың түрлі анықтамаларын қатар қойып қараңыз. Сіздің ойыңызша, қай анықтама лингвистикаға сәйкес келеді?
3. Әртүрлі дереккөздерінде ерекшеленген ақпарат қасиеттерін салыстырып, ақпаратқа қатысты қасиеттердің жіктемесін толық анықтаңыз.
4. Ақпарат, хабарлама, мәліметтер сәйкестілігін түсіндіріңіз.
5. Ақпараттық технологиялардың негізгі даму кезеңдерін анықтап атаңыз.
6. Зерттеушілер ақпараттық технологиялардың болашағын неден көреді? Осыған байланысты сіздің ойыңыз қандай?
7. Міндет пен ереже дегеніміз не? Бұл ұғымдардың алгоритмге байланысы қандай? Алгоритімнің қасиеттері қандай?

### Әдебиеттер:

#### *Негізгі:*

1. Основы теории коммуникации: учебник / под ред. М.А. Василика. – М.: Гардарики, 2007.
2. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
3. Большакова Е.И. Компьютерная лингвистика: методы, ресурсы, приложения // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 90-105.
4. Алексеев В. Услышь меня, машина // Компьютера. 1997. №49. <http://offline.computerra.ru/1997/226/938> (қаралған күн: 28.02.2012).
5. Степанов А.Н. Информатика: учеб. пособие. – СПб.: Питер, 2006.
6. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
7. Ожегов С.И. Словарь русского языка. 22-е изд., стер. – М.: Русский язык, 1990.
8. Гейн А.Г., Сенокосов А.И. Справочник по информатике для школьников. – Екатеринбург: У-Фактория, 2003.
9. Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998.

#### *Қосымша:*

1. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд., испр. – М.: Флинта: Наука, 2007. – С. 9-16.



### **3-дәріс. Лингвистикадағы ақпараттық технологиялардың аппараттық және бағдарламалық қамтамасыз етілуі**

**Қарастырылатын мәселелер:** *компьютер мен периферийлік/шеткі құралдар – ақпараттық технологиялардың аппараттық негізі; жүйелік пен қолданбалы бағдарламалық қамтамасыз етілу; лингвистикалық ресурстар (lingware); лингвистің автоматтандырылған жұмыс орны.*

**Тірек сөздер мен сөз тіркестері:** *компьютер, компьютерлік бағдарлама, бағдарламалық қамтамасыз етілу, жүйелік бағдарламалар, автоматты өңдеу, компьютерлік ресурстар, лингвистің автоматты жұмыс орны, қолданбалы бағдарламалар.*

*Компьютерлік бағдарлама* (ағыл. *computer program*) – 1) белгілі бір алгоритмді жүзеге асыру мақсатында мәліметтер өңдеу жүйесінің нақты құрауыштарын (компьютерді) басқаруға арналған мәліметтер тізбегі; 2) орындалуға (өңделуге) тиіс реттелген бұйрықтар тізбегі, есеп шығару алгоритмін сипаттайтын бағдарламалау тілінің сөйлемдер жиыны. Есеп шығаруға, сондай-ақ берілген мәселені шешуге арналған, қабылданған синтаксиске сәйкес жазылған компьютер бұйрықтарының (нұсқауларының) реттелген тізбегі; 3) бағдарламалар мен олардың үзінділеріне, бағыныңқы бағдарламаларға, процедураларға, бағдарламалық модульдерге, машиналық бағдарламалар мен олардың үзінділеріне арналған жалпы атау.

ЭЕМ жұмысындағы бағдарламалар және оның іс-әрекетінің реттілігі мен ережелері, оған қосылған құрылғылардың да қызмет ету ерекшеліктерін анықтауға жол ашады. Әрбір компьютердің бағдарламалар жиыны оны пайдаланудың бар мүмкіндіктерін анықтап береді. Егер бағдарламалары болмаса, машина ешқандай да жұмыс атқара алмайды. Кез келген ЭЕМ-мен жұмыс істеу оның бағдарламаларын оқып үйренуден басталады. Ал әрбір үлкен, орташа, шағын немесе дербес ЭЕМ түрлерінде көптеген бағдарламалық жабдықтар бар, сондықтан ЭЕМ-ді үйрену оның бағдарламаларымен танысудан басталады.

Лингвистикалық мәліметтерге көлемді есеп жүргізуге, сонымен бірге лингвистикалық модельдеуді орындау үшін ЭЕМ-ді

(немесе компьютер) қолдану тиімді. *Компьютер* – адам құрастырған алгоритм (немесе бағдарлама) бойынша машинаға түсінікті тілде жазылған ақпаратты автоматты құру, өңдеу, жеткізу, іске қосу үшін жұмыс істейтін электронды құрылғы [1, 42-б.; 15, 22-б.].

Келтірілген анықтама бойынша, компьютерді қолдануда ақпараттық технологиялардың аппараттық (*hardware*) және бағдарламалық қамтамасыз етілуі (*software*) қатар жүреді.

Аппараттық қамтамасыз етілуге компьютердің өзі (станционарлы немесе көшірілмелі), сонымен қатар қолданушы арқылы компьютерге ақпаратты енгізу/шығаруға (*пернетақта, маус, монитор, принтер* және т.б.) немесе компьютерді өзге құрылғылармен байланыстыратын (мысалы, модем) перифериялық құралдар жатады.

*Бағдарламалық қамтамасыз етілу* – аппараттық құралдарды басқаруға бағытталған немесе ақпаратқа, сәйкес құжаттарға қатысты әртүрлі операцияларды орындауға арналған машина тілінде жазылған бұйрықтардың бірізділігін білдіретін компьютерлік бағдарламалар.

Бағдарламалық құралдардың тағайындалуына байланысты бағдарламалық қамтамасыз етілу жүйелік және қолданбалы болып қарастырылады. *Жүйелік бағдарламалар* аппараттық құралдардың жұмыстарын басқарады әрі операциялық жүйелер, утилиттер, драйверлер және басқа да бағдарламалардың түрлерін құрайды. *Қолданбалы бағдарламалар* соңғы қолданушыға арналған және ақпарат бойынша түрлі операцияларды орындауға мүмкіндік береді:

- мәтінді құру мен өңдеу (мәтіндік редакторлар);
- графикалық көріністерді өңдеу (графикалық өңдеу);
- дыбыстық және видеоақпаратпен жұмыс (мультимедиялық бағдарламалар);
- статистикалық мәліметтерді өңдеуге арналған электронды кестелерді құру (электронды кестелер) және т.б.

Лингвист үшін электронды аудармашылар мен сөздіктер, сондай-ақ мультимедиялық оқытушы бағдарламалар сияқты қолданбалы бағдарламалар өте тиімді.

Кейбір зерттеушілер ақпараттық технологиялардың аппараттық және бағдарламалық қамтамасыз етілуімен (БҚ) бірге барлық лингвистикалық компьютерлік ресурстарды (грамматикалық анықтамалар, сөздіктер, энциклопедиялар, мәліметтердің лингвис-

тикалық қоры және т.б.) білдіретін *lingware* (немесе *linguware*) ұғымын қолданады [2, 27-б., 31, 59-б.].

Лингвистикалық мәліметтерді автоматты өңдеуге қажетті аппараттық, бағдарламалық және лингвистикалық құралдарды *лингвистің автоматты жұмыс орны* (АЖО) дейміз [3, 258]. Лингвистің АЖО-сы компьютердің өзін, операциялық пен базалық қолданбалы (БК), сонымен қатар ана тілі мен оқытылатын шет тілдеріне қатысты барлық лингвистикалық компьютерлік ресурстарды қамтиды.

Лингвистің АЖО-сы мамандандырылуына сәйкес шет тілін аудару немесе оқытуға байланысты қолданбалы бағдарламалар және лингвистикалық ресурстармен толықтырылуы мүмкін. Білім алушылардың міндеті – аппараттық және бағдарламалық қамтамасыз етілудің қазіргі қалпын жетілдіруді қамтитын АЖО-ны тұрақты дамыту, сондай-ақ өзіндік лингвистикалық ресурстық қордың, яғни лингвистикалық анықтамалар, сөздіктер, деректер қорларын іздеу, сақтау, алу немесе құрумен тұрақты толықтырып отыру.

### Семинар тапсырмалары

1. Компьютер құрылысын сипаттап, перифериялық/шеткі құралдарды көрсетіңіз.
2. Жүйелік пен қолданбалы бағдарламалық қамтамасыз етілуге анықтама беріңіз. *Операциялық жүйе, утилит, драйвер* ұғымдарын түсіндіріңіз.
3. Қолданбалы компьютерлік бағдарламалардың топтастырылуына тоқталыңыз. Оларды қысқаша сипаттап, қолданбалы компьютерлік бағдарламалардың негізгі түрлеріне мысалдар келтіріңіз.
4. Мәтіндік процессорды және оның лингвистикалық қызметтерін сипаттаңыз.
5. Лингвистикалық мақсаттарға арналып құрастырылған арнайы компьютерлік бағдарламаларға тоқталыңыз.
6. Компьютерлік лингвистиканың лингвистикалық ресурстары туралы айтыңыз (*lingware*).

### Әдебиеттер:

#### *Негізгі:*

1. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие для студ., аспирантов, преподавателей филологических факультетов. 2-е изд. – М.: Флинта: Наука, 2007.
2. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных

- технологиях: учеб. пособие. – СПб.: Книжный Дом, 2007.
3. Rayner M., Carter D. M., Bretan I., Eklund R., Wirén M., Hansen S.L., Kirchmeier-Andersen S., Philp C., Sorensen F., Erdman Thomsen H. Recycling Lingware in a Multilingual MT System // *Computation and Language*. 1997. [www.aclweb.org/anthology/W/W97/W97-0910.pdf](http://www.aclweb.org/anthology/W/W97/W97-0910.pdf) (каралган күн: 28.02.2012).
  4. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. М.: Академия, 2006.
  5. Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998.

*Қосымша:*

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 97-99.
2. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд. – М.: Флинта: Наука, 2007. – С. 22-26, 37-53.
3. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие для студ. вузов. – М.: Академия, 2004. – С. 19-22.
4. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации: учебно-метод. пособие. – М., 2009. – С. 92-102.
5. Степанов А.Н. Информатика: учеб. пособие. – СПб.: Питер, 2006. – С. 42-43, 80-85, 10-111.
6. Чухарев Е.М. Компьютерные технологии в лингвистических исследованиях: указания по выполнению домашнего задания. – Архангельск, 2009. – С. 2-4.

# АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР: ЛИНГВИСТИКАДА ҚОЛДАНЫЛУ АЯСЫ

### **4-дәріс. Дыбыстық сөйлеудің автоматты анализі мен синтезі**

**Қарастырылатын мәселелер:** *сөйлеуді автоматты талдаудың кезеңдері; дыбыстық сөйлеуді компьютерге енгізу; аналогты және сандық сигнал; сөйлеуді сөзбе-сөз және фонемалық талдау; дыбыстық сөйлеу мен компьютерді дауыспен басқаруды өңдеу бағдарламалары; сөйлеуді автоматты синтездеу әдістері.*

**Тірек сөздер мен сөз тіркестері:** *дигитализациялау, дыбыстық сөйлеу, жазбаша мәтін, өңдеу жұмыстары, автоматты өңдеу, синтездеу, автоматты талдау, дыбыстық сигнал, сигналдар аналогы, таңба, сөйлеуші тұлғасы, сөздік қоры, сөйлеудің дайындық деңгейі, акустикалық сигналдар, сөйлеудің автоматты синтезі, интерфейс.*

Ақпараттық технологияларды лингвистикада қолданудың алғашқы маңызды қадамдарының бірі мәтінді **дигитализациялау** – жазбаша және ауызша формадағы тілдік материалды сандық қалыпқа түсіру болып табылады. Тек дигитализациялау кезінде ғана табиғи тілдегі мәтіндерге арнайы әрекеттер жасау үшін компьютер көмегіне жүгіну мүмкін болады. Мәселен, табиғи тілдегі

мәтіндерді өңдеу, оның жеке элементтерін анықтау және оған сай мәтіндерді құрастыру.

Дыбыстық сөйлеу және жазбаша мәтіндерді дигитализациялау мен өңдеу жұмыстары өзгеше болғандықтан, біздің еңбегімізде бұл мәселелер әртүрлі тарауларда қаралады. Алғашқы тарау дыбыстық сөйлеуді автоматты өңдеу және синтездеу мәселелеріне арналса, қалған тараулар жазбаша мәтіндерді автоматты өңдеуге арналады.

Дыбыстық сөйлеуді автоматты түрде талдау барысында ол бірден жазбаша қалыпқа ауысады, себебі жазбаша мәтін қажетті операцияларды әрі қарай жалғастыруда қолайлы болып есептеледі. Дыбыстық сөйлеуді автоматты талдау дегеніміз – сандық формадағы жазбаша мәтінді табиғи тілдегі дыбыстық мәтінге айналдыру.

Мәтінді автоматты талдау үдерісі төмендегі кезеңдерден тұрады:

- микрофон көмегімен дыбыстық сөйлеуді компьютерге енгізу;
- дыбыстық лектегі жекелеген таңбаларды компьютерлік бағдарламалар көмегімен ерекшелеу;
- дыбыстық сөйлеуде анықталған таңбаларды тілдің таңбаларымен сәйкестендіру.

Дыбыстық сөйлеудің ең кіші таңбасы ретінде адамның артикуляторлық аппаратымен шығарылатын дыбыстар саналады. Әр дыбыстың арнайы құралдармен (мәселен, осцилографпен) өлшеуге болатын қасиеттері (дыбыс жоғарылығы, дыбыс толқындарының жиілігі) бар.

Дыбыстық сигналдың параметрлері үздіксіз өзгеріп отырады. Ондай *сигналдар аналогы (сигналдық аналогы)* деп аталады. Аналогты сигналдардан айырмашылығы сандық сигналдар дыбыстық толқынның әртүрлі деңгейлерін қадағалауға болатын жеке сандық мәліметтерден тұрады. Микрофонды пайдалану барысында аналогты дыбыстық сигнал заманауи компьютерлердің дыбыстық картасына жүктелген аналогты-сандық өзгерткіштер арқылы дискретті сандық сигналға ауысатын аналогты электрлі сигналдарға өзгереді [49].

Бүгінгі күні өте көп болып саналатын ауызша сөйлеуді автоматты анықтау құралдары алғашында сөйлеу легінде ерекшеленетін таңба ретінде дыбыс емес, сөзді есептеген. Компьютерге енгізілетін сөйлеулер диктор алдын ала жазып алған сөздермен

сәйкестендіріледі. Бірақ, сөйлеуді анықтаудың бұл түрі белгілі бір шектеулермен байланысты:

– сөйлеуші тұлғасы: автомат тек белгілі тұлғаның сөйленісін анықтай алады;

– сөздік қоры: автомат шектеулі сөздерді ғана анықтайды,

– сөйлеудің дайындық деңгейі: автомат тек алдын ала дайындалған сөйленісті анықтай алады [1, 39-б.].

Бұл шектеулерді жеңу үшін компьютерлік бағдарламалар сөзді емес, дыбысты анықтай алуы керек, бұл компьютер үзілістері бар дискретті сөйлеуді емес, біріккен табиғи сөйлеуді анықтай алу қажет дегенді білдіреді.

Сөйлеу дыбыстарын фонема бойынша талдау негізінде: дыбыс-талудың ұзақтығы мен жылдамдығы, акустикалық сигнал мен үзілістерді талдау жатыр.

Қазіргі таңда қазақстандық және шетелдік тіл білімінде интонацияның табиғатын анықтауға арналған теориялық материалдар баршылық екені бәрімізге белгілі. Осы құбылысты зерттеу барысында интонацияның кең және тар мағынада қолданыс табылған екі түрлі түсініктемесі анықталды, атап айтсақ, кең (кешенді, көпкомпонентті) мағынада интонация гетерогенді просодикалық құбылыстардың (интономалардың) комбинациясы ретінде анықталды, яғни әр параметрдің өзіндік бірліктері мен мағыналары бар (тономалар, хронемалар, акцентемалар және т.б.) просодикалық құбылыстарға негізгі тон, темп, түйіскен жер типтері, интенсивтілік, ритмикалық құрылымдарды жатқызады. Бұл бағыттың өкілдері: А.А. Абдуазизов, В.А. Артемов, Ш. Атенов, З.М. Базарбаева, И.А. Баймұратова, В.А. Богородицкий, В.А. Васильев, Д. Джоунз, В.Н. Всеволодский-Гернгросс, Ж.К. Қалиев, М.Г. Қаспарова, Ж.А. Аралбаев, Т. Кеншинбаев, Дж. О'Коннор, Д. Кристал, Б.К. Мұрзалина, А.Н. Нұрмаханова, А.М. Пешковский, Г.П. Торсуев, Н.У. Туркбенбаев, Ж.М. Утесбаева, А.М. Фазылжанова, Л.К. Цеплитис және т.б. Тар мағынада, интонация – бұл тек тономалардан тұратын мелодика, яғни Қ. Жұбанов, Ә. Хасенов, Р. Сыздық айтқандай, айтылу/ дауыс/сөз сазы және қазіргі кезде жиі пайдаланып жүрген әуен деген атау да осы құбылысқа сай келетінін баса айтпақпыз. Осы пікірді ұстанған ғалымдар: С. Аманжолов, Л. Армстронг, А. Байтұрсынов, М. Балақаев, Н.А. Баскаков,

Д. Болинджер, К. Жұбанов, Ә. Жүнісбек, І. Кеңесбаев, С.В. Кодзасов, М.И. Матусевич, К. Пайк, Г.Е. Палмер, С. Оде, И. Уорд, Л.В. Щерба және т.б.

Шетелдік еңбектерде интонацияны сипаттаудың басқа да жолдары бар. Интонация практикалық емес құбылыс ретінде танылып, оның компоненттері ретінде тек интенсивтілік, ұзақтылық және тембрді ғана жатқызған. Осындай мектептердің бірі Голланд мектебі, яғни интонацияны сөйлеу әуені (тоналды акцент) ретінде анықтайды [2, 9-б.; 3, 218-б.] және осы құбылысты сипаттауда жиі эксперименталды әдісті жүзеге асыру үшін компьютерлі бағдарламаларды қолдана отырып зерттеуде. Әуен немесе мелодика интонацияның ең маңызды компоненті ретінде танылған, себебі сөйленімнің әуендік безендірілуіндегі әртүрлілігі оның қабылдауы мен интерпретациясын анықтайды: сөйлеуші коммуникативті маңызды мәліметті жеткізу үшін сөйленімдегі керек фрагменттерді (сөздерді, синтагмаларды) бөліп көрсетеді, мысалы, тон жоғарылығын көтерілу немесе түсіру арқылы тоналды акцентті жүзеге асырады.

Бірінші болып тоналды акцент терминін Д. Болинджер қолданды, сонымен тоналды акцент дегеніміз: «Егер буындағы акцент тон жоғарылығы арқылы жүзеге асса, онда тон жоғарылығы екі қызмет атқарады. Біріншіден, ол осы буындағы акцент туралы белгі береді. Екіншіден, оның жоғары, төмен немесе оның бағыттарының деңгейі әуен жасайды. Осы контурдың қозғалысы белгілі бір әуенді жүзеге асырады. С. Оде, Д. Болинджердің тоналды акцентін қарастыра отырып, мынаған зер салған: «Тоналды акценттің жүзеге асуы тек айналасындағы сөздердің ішіндегі белгілі бір маңызды сөзді бөліп көрсету ғана емес, сонымен қатар сөйлеу әуенін де жүзеге асырумен де көрінеді» [4].

С. Оде бойынша, «интонация – бұл сөйлеу ағымындағы тон жоғарылығының өзгерісі. Тон жоғарылығының вариациясы – бұл тон қозғалысы, жоғары, төмен, түзу, тон жоғарылығы диапазонындағы жоғарыға немесе төменге секіріс, сондай да басқа тонның қозғалыстары да көрініс тапқан. Тонның белгілі бір қимылы немесе тонның конфигурациясы сөздегі буын перцептивті проминенттілікке ие болады. Егер проминенттілік тон жоғарылығы арқылы жүзеге асырылса, онда осы буын сөз екіні бар тоналды акцентке ие болады. Сонымен айналадағы сөздермен салыстырғанда сөз тоналды



акцент арқылы ерекше бөлініп көрсетіліп, үлкен проминенттілікке ие болады. Сөйлеуші белгілі бір сөздерде тоналды акцентті жүзеге асырып, тыңдаушыға сөздегі маңызды мәлімет ұстанған сөз туралы хабарлайды, сөздің басқа сөздермен байланысын, көңіл-күйді білдіретін сөздерді көрсетеді. Бұл қызметті тек тоналды акцент атқарады деп айта алмаймыз. Сонымен бірге басқа да просодикалық параметрлер, мысалы, интенсивтілік немесе дауыстың ұзақтылығы, сөйлеу қарқынындағы вариация мен ритмика сөзге немесе сөздер тобына проминенттілік бере алады» [4, 210-б.; 5, 6-б.].

Қазақ тілінің интонациясын зерттеуге алғаш рет сөйлеу сигналдарын өңдеуге және талдауға арналған Praat, Doing phonetics by computer атты компьютерлік бағдарлама қолданыс тапты. Praat бағдарламасын қолдана отырып, дикторларға алдын ала дайындалып берілген қазақ тілінің хабарлы сөйленімдері жазылды, қазақ тілінің хабарлы сөйленімдерінің интонациясы фонетикалық әдіс арқылы зерттелді. Бұл бағдарламаны сегменттік және суперсегменттік деңгейдегі кез келген тілдердің дыбыстық жүйесін зерттеуге пайдалануға болады. Сонымен қатар осы бағдарлама көмегімен диссертациялық зерттеу негізінде дайындалған және дайындықсыз жазылған қазақ тілінің интонациялық бірлігін сипаттауға тырысты [6].

Амстердам университетінің Фонетика институтының қызметкерлері Paul Boersma и David Weenink (2006) өңдеп, ойлап тапқан Praat бағдарламасының ыңғайлылығы ең соңғы нұсқасын интернеттен сақтап алу мүмкіндігі болып табылады (нидерланды; [www.fon.hum.uva.nl/praat](http://www.fon.hum.uva.nl/praat)). Praat бағдарламасының беретін мүмкіндіктері мынадай:

– Praat сөйлеу сигналын зерттеуге, өлшеуге, модификациялауға және өңдеуге арналған, ол сонымен бірге кез келген табиғи тілдің тон жоғарылығын өлшеуге және интонациясын зерттеуге мүмкіндік береді;

– Praat бағдарламасын интернеттен тегін сақтап алуға болады;

– Praat бағдарламасы компьютердің кез келген түрінде жұмыс істейді;

– Praat бағдарламасын ғылыми мақсатта және оқу үдерісінде пайдалануға болады.

Тоналды акценттердің формасын сипаттау үшін тон жоғарылығы өлшеніп, акценттелген фрагментте көрініс тапқан тон қи-

мылының ең жоғарғы және ең төменгі нүктесі арасындағы аралық көлемі анықталды. Осылай тоналды акцент бір немесе бірнеше тон қимылынан тұратындығы айқындалды.

Аралық (интервал) арақашықтығының логарифмдік шәкілдегі бірлігі ретінде «полутон» алынды (ағылшын. semitone, қысқартылғаны ST). Тон қозғалысы герц арқылы өлшенді; аралық көлемі полутон арқылы өлшемін тапты, себебі адамзаттың есту қабілеті тон жоғарылығы сызықты шәкілде емес, логарифмдік шәкілде қабылданады.

Сонымен, біз мелодикалық аралық арасындағы арақашықтықты полутон арқылы сипаттап, логарифмдік шәкілдегі Praat бағдарламасының көмегімен контурлардың суреттерін саламыз.

Әртүрлі регистрлі дикторлардың аралығын салыстыру аралықты өлшеуге арналған полутон арқылы өлшенген өлшемдер сипаттауды жеңілдетеді. Мысалы, аралық қабылдау көлемі 12 полутонға тең әйел регистрі 12 полутонға тең еркек регистріне сәйкес келеді. Ал аралық қабылдауы әйел регистріндегі (мысалы, 200-ден 300 гц-ке дейін) 100 гц тең еркек регистріндегі (мысалы, 100-ден 200 гц-ке дейін) аралық көлеміне сәйкес келмейді (әртүрлі шәкіл сипаттау тәжірибесінен қараңыз. Оде) [5].

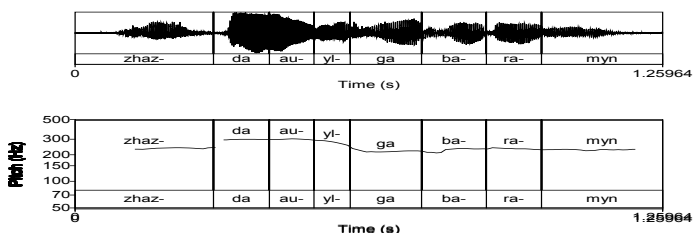
Сонымен, әр сөйленімге буын арасындағы арақашықтықты көрнекі көрсететін мәтінді суреттер жасалынған. Барлық сөйленімдердің мәтіндері суретке енгізілген, сондықтан әр суреттегі интонациялық контурлардың әр сөйленім сегментінің орнын анық көруге болады (төмендегі суреттерді қараңыз).

Аталмыш бағдарлама арқылы жазылған қазақ тіліндегі хабарлы сөйленімдердің талдауын бағдарлай аласыздар. Бұл зерттеудің мақсаты – жазылған сөйленімдердегі тоналды акценттің орнын, қызметін анықтап, адресат пен адресант арасында атқарып тұрған қызметін сипаттау болып табылады.

Хабарлы сөйленімдердің кезекті талқылауына назар аударайық.

1. Жазда	ауылға	барамын.
zhazda	auylǵa	baramyn
Летом	в ауыл	поеду

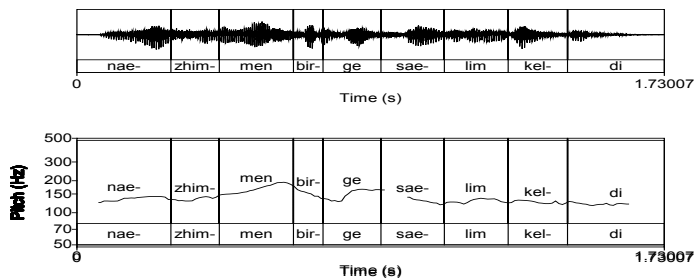
Бұл сөйленімде D2 дикторы айтқан жаз-да сөзінің соңғы буынында өрлімелі тон анықталған. D2 үшін маңыздысы жазда деген сөз болып тұр, яғни ауылға басқа жыл мезгілі емес жазда барады.



**1-сурет.** Әйел адам айтылымындағы жазда ауылға барамын хабарлы сөйленімі. (D2-15) әйел дауысымен айтылған хабарлы сөйленім. Жоғарғы сурет – сөйленімнің осциллограммасы. Тік сызықтар буындар арасындағы шекараларды көрсетіп тұр. Төменгі суреттегі көлденеңнен уақыт секундпен, ал тігінен логарифмдік шәкілдегі негізгі тон жиілігі герц арқылы берілген.

-Да буынындағы жоғарылау аралығының өлшемі 247 гц-тен 307 гц-ке шейін; полутон өлшемімен алғанда екі нүкте арасындағы аралық 4ST-ге тең, бұл сөйленім дәстүрлі зерттелген қазақ тіліндегі хабарлы сөйленіммен сәйкес келмейді. Бұл мысал әрқашанда хабарлы сөйленім соңында төмен түсетін тонмен аяқталады деген тұжырымға қайшы келіп тұр, сонымен қатар оған түзу тон жоғарылығы тән және осы орташа деңгейде қалып қояды деген болжамды дәйектеп тұр. Ал өрлімелі тоналды акценттің орны барлық хабарлы сөйленімдермен сәйкес екенін көрсетіп тұр.

2. Нәжіммен            бірге            Сәлім            келді.  
 nāzhimmen            birge            sālim            keldi  
 с Нажимом            вместе            Салим            пришел  
 Вместе с Нажимом пришел Салим.



**2-сурет.** Еркек адам айтылымындағы Нәжіммен бірге Сәлім келді – хабарлы сөйленімі. (D6-41) еркек дауысымен айтылған – хабарлы сөйленім. Жоғарғы

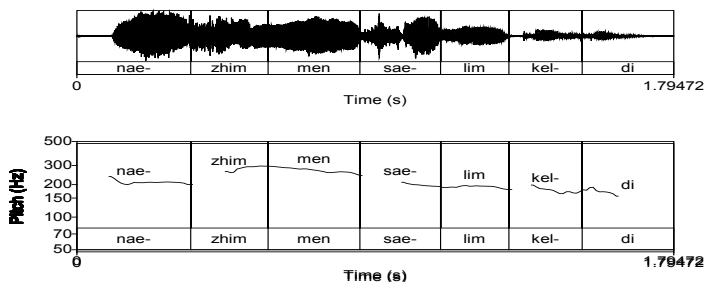
ғы сурет – сөйленімнің осциллограммасы. Тік сызықтар буындар арасындағы шекараларды көрсетіп тұр. Төменгі суреттегі көлденеңнен уақыт секундпен, ал тігінен логарифмдік шәкілдегі негізгі тон жиілігі герц арқылы берілген.

Д6 Нә-жім-мен деген сөздің соңғы буынында өрлімелі тоналды акцентті жүзеге асырды. Д6 үшін Нәжіммен бірге Сәлім келгені маңызды.

-Мен буынындағы интервал жоғарылауы 155 гц-тен 201 гц-ке тең; полутонмен алғанда екі нүкте арасындағы аралық өлшемі – 5ST. Осы мысалдар өрлімелі тоналды акцент хабарлы сөйленім басынан орын алып, ал сөйленімнің өзі төмен түсетін немесе түзу тон жоғарылығымен аяқталатынын көрсетіп тұр.

3. Нәжім	мен	Сәлім	келді.
nāzhīm	men	sālīm	keldī
Нажим	и	Салим	пришли

Нажим и Салим пришли.



**3-сурет.** Әйел адам айтылымындағы Нәжім мен Сәлім келді – хабарлы сөйленімі. (D3-42) – әйел дауысымен айтылған хабарлы сөйленім. Жоғарғы сурет – сөйленімнің осциллограммасы. Тік сызықтар буындар арасындағы шекараларды көрсетіп тұр. Төменгі суреттегі көлденеңнен уақыт секундпен, ал тігінен логарифмдік шәкілдегі негізгі тон жиілігі герц арқылы берілген.

Екі түрлі диктор, яғни Д6 (2-сурет Нәжіммен бірге Сәлім келді) және Д3 (3- сурет Нәжім мен Сәлім келді), арқылы айтылған сөйленімдердегі тоналды акценттің орындалуын салыстырайық:

– екі сөйленімде де бірдей екі ақиқат бар: Екі адам келді. Бірінші сөйленімде Д6 тоналды акцентті бірге деген сөзде орындады,

ол үшін Сәлім Нәжіммен бірге келгені аса маңызды. Екінші сөйленімде керісінше, басқа мәліметті көрсеткісі келген, яғни Нәжім мен Сәлім келді. Бірінші сөйленімде жасырын ақпарат бар, мүмкін Сәлім осы жерге келгісі келмегендіктен, оны Нәжім өзімен бірге ертіп келген шығар. Ал екінші сөйленімде екеуі қатар келді, бұл жерде ешқандай жасырын ақпарат жоқ, бұл жерде екі кейіпкер бір-біріне тәуелсіз, байланыс байқалмайды, айтушы екеуі де келді деген ақпарат беріп отыр. Тоналды акценттің орындалуы маңызды ақпараттың қайда екендігі жайлы хабарлап тұр десек те болады.

– екі сөйленімде де тоналды акцент Нәжім-мен деген сөздің соңғы буынынан орын тапқан, -мен буынындағы аралықтың жоғарылауы 155 гц-тен 201 гц-ке тең; полутонмен екі нүкте арасындағы аралық өлшемі – 5ST. D3-те Нәжім сөзінің - жім буынындағы аралық жоғарылауы 209 гц-тен 309 гц-ке тең; полутонмен екі нүкте арасындағы аралық өлшемі – 7ST. Екі сөйленімнің сипаттамасы дәстүрлі қазақ тілінің хабарлы сөйленімдерінің мысалы ретінде қарастырыла алады.

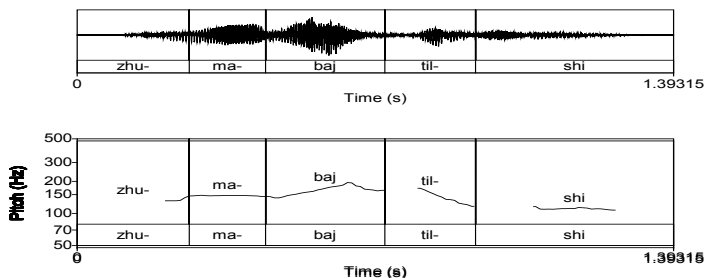
– екі сөйленімде де өрлімелі тоналды акцент сөйленім басында орындалған, D6-ның орындауында ол түзу сызық болса, D3-тің орындауында оған сөйлем соңында хабарлы сөйленімдердің жалпы суреттеріне тән интонологиялық сызық төмен түскен.

4. Жұмабай –	тілші.
zhūmabai	tilshī
Жумабай –	журналист.

D6 бұл сөйленімді екі синтагмаға бөлген: Жұмабай/ тілші//:

– Жұмабай бірінші синтагмасында Жұ-ма-бай соңғы буынында өрлімелі тоналды акцент орындалған. -бай буынындағы аралық жоғарылауының өлшемі 146 гц-тен 203 гц-ке тең. Полутонмен екі нүкте арасындағы аралық өлшемі – 6ST.

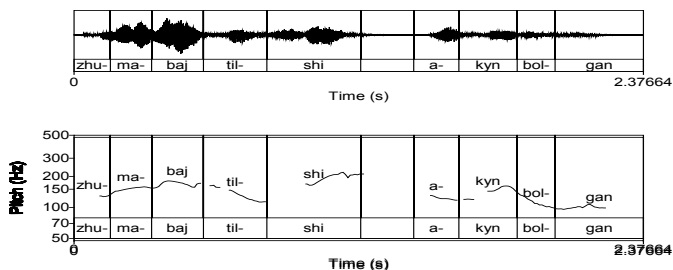
– Екінші синтагманың тіл-ші сөзінің бірінші буынында қосымша төмен түскен тоналды акцент орындалған. Тіл буынындағы аралық төмен түсуінің өлшемі 180 гц-тен 122 гц-ке тең. Полутонмен екі нүкте арасындағы аралық өлшемі – 7ST.



**4-сурет.** Еркек адам айтылымындағы Жұмабай – тілші хабарлы сөйленімі. (D6-45) еркек дауысымен айтылған хабарлы сөйленім. Жоғарғы сурет –сөйленімнің осциллограммасы. Тік сызықтар буындар арасындағы шекараларды көрсетіп тұр. Төменгі суреттегі көлденеңнен уақыт секундпен, ал тігінен логарифмдік шәкілдегі негізгі тон жиілігі герц арқылы берілген.

D6 орындауымен жасалған Жұмабай – тілші деген сөйленім Жұмабай, тілші, ақын болған деген сөйленімнің құрамдас бөлігі ретінде интонациялық суреті күрт өзгерген. Бұл сөйленімді хабарлы сөйленім ретінде алсақ, қазақ тілінің дәстүрлі сөйленімдерімен сәйкес келеді.

5. Жұмабай,	тілші,	ақын	болған.
zhūmabaj	tilshī	aқyn	bolğan
Жумабай	журналист	поэт	был
Жумабай	был	журналистом,	и поэтом.



**5-сурет.** Еркек адам айтылымындағы Жұмабай, тілші, ақын болған –хабарлы сөйленімі. (D6-46) – еркек дауысымен айтылған хабарлы сөйленім. Жоғарғы сурет – сөйленімнің осциллограммасы. Тік сызықтар буындар арасындағы шекараларды көрсетіп тұр. Төменгі суреттегі көлденеңнен уақыт секундпен, ал тігінен логарифмдік шәкілдегі негізгі тон жиілігі герц арқылы берілген.

Д6 сөйленімді мынадай екі синтагмаға бөледі: Жұмабай тілші, ақын болған//. Аталмыш сөйленімде екі өрлімелі тоналды акцент орындалған: Жұ-ма-бай және тіл-ші.

Тіл-ші сөзінің соңғы буынында маңызды тоналды акцент орындалған, себебі сөйленім аяқталмаған және оны айтушы аяқталмаған интонация арқылы білдіріп тұр, яғни өрлімелі тоналды акцент арқылы орындаған, Жұмабай тілші бірінші сөйленіммен салыстырғанда, ол бірінші буынды емес тіл-ші деген екінші буынды бөліп айтқан. -Бай буынындағы аралық жоғарылауының өлшемі 165 гц-тен 189 гц-ке тең. Полутонмен екі нүкте арасындағы аралық өлшемі – 3ST. -Ші буынындағы аралық жоғарылауының өлшемі 118 гц-тен 228 гц-ке тең. Полутонмен екі нүкте арасындағы аралық өлшемі – 12ST.

Д6 айтқан Жұмабай – тілші сөйленімді осы диктор айтқан сөйленіммен Жұмабай, тілші, ақын болған деп салыстырайық.

Д6: Жұмабай – тілші.

– Бірінші синтагмадағы Жұ-ма-бай сөзінің буынындағы полутонмен аралық өлшемі – 6ST. Екінші синтагмадағы тіл-ші сөзінің буынындағы екі нүкте арасындағы аралық өлшемі полутонмен – 7ST.

– Жұмабай аяқталмаған интонация арқылы орындалды.

– Жұмабай – тілші аяқталған интонация арқылы орындалды.

– Сөйленім соңы төмен түскен тонмен сипатталады.

Д6: Жұмабай, тілші, ақын болған.

– Бірінші синтагмада Жұ-ма-бай буынындағы полутонмен екі нүкте арасындағы аралық жоғарылауының өлшемі – 3ST. тіл-ші буынындағы полутонмен алғанда екі нүкте арасындағы аралық жоғарылауының өлшемі – 12ST.

– Жұмабай аяқталмаған және атап шығу интонациясымен орындалған.

– Жұмабай – тілші аяқталмаған және атап шығу интонациясымен орындалған.

– Сөйленім төмен түскен тонмен аяқалған.

Қазақ тілінің хабарлы сөйленімдеріндегі тоналды акценттерді зерттеу барысында мынадай қорытындыға келдік:

– хабарлы сөйленімнің басы өрлімелі тоналды акценттен (бірінші сөздің соңғы буынында) және сөйленім соңы төмен түсетін немесе түзу тон жоғарылығы арқылы жасалған;

– өрлімелі тоналды акценттің аралық өлшемі 1ST-ден 3ST-ге дейін (минималды өлшем), 4ST-ден 6ST-ге дейін (орташа өлшем), 7ST-ден 12ST-ге дейін (максималды өлшем) диапазонында байқалды. Аралық өлшемі маңызды ақпарат ұстанған тоналды акценттің орны жайлы хабарлап отырды;

– дайындықпен жазылған хабарлы сөйленімдердегі тоналды акценттерді анықтау барысында айтарлықтай маңызды ерекшеліктер байқалмады.

Дыбыстық сөйлеуді тануда оның барлық қасиеттері мен ерекшеліктері машина жадындағы сөздің үлгілері және қасиеттерімен салыстырылады. Салыстыру нәтижесінде сөз транскрипцияланады және сөздің орфографиялық жазбасы жасалынады.

Біріккен сөйлеуді автоматты түрде талдауда сөздер арасында шекараның болмауы біршама қиындық тудырады. Бұл қиындықты шешу үшін адам акустикалық сигналдармен қоса, басқа да ақпарат көздерін пайдаланады: белгілі бір жағдайды я оқиғаны, мәнмәтінді, тілдік айтылымның құрылымын, осы сала бойынша өткен тәжірибесін, т.с.с. Осы іспеттес заңдылықтарды ғалымдар машиналарға да қолданып, сөйлеуді талдаудың қазіргі жүйелеріне акустикалықтан өзге лексикалық, синтаксистік, семантикалық, прагматикалық сияқты тіл жүйесінің деңгейлерін қатыстыруды көздейді.

Сөйлеуді автоматты талдауға семантикалық деңгейді қосу төмендегідей салдарлар тудырады:

– машина енгізілген сөздерді көпмағыналы және шынайы деп таниды;

– машина кейбір сөйлеу мәнмәтіндерінде қарым-қатынастың белгілі бір түрлері пайда болуын болжайды; қарым-қатынастың қай түрін болса да болжау арқылы машина сөйлемді талдайды, интерпретациялайды [6,120-б.].

Бұл жерде сөйлеуді автоматты талдаудың мұндай күрделі деңгейі әртүрлі мамандық өкілдерінің қатысуын талап етеді. Ғалымдар мен тәжірибе жасаушылардың уақыты мен күшін үнемдеу үшін көп компаниялар, мәселен, Microsoft, сөйлеуді талдау құралдарын бағдарламалық модульдер және интерфейстер ретінде шығарады.

Лингвистика, математика, биология салалары бойынша білімі жоқ программистер өз жұмыстарында дайын интерфейстер мен модульдерді пайдалана алады. Бірақ бұл кезде бағдарламалардың сөйлеу мүмкіндіктері қолданылған құралдар және технологияларымен шектеледі. Мысалы, егер сөйлеуді анализ және синтездеу құралдары орыс тілімен жұмыс істей алмаса, олар Ресейде пайдаланылмайды [7].



Сөйлеуді автоматты талдау құралдары қолданылатын төмендегі бағдарламаларды атасақ болады:

- компьютер және тұрмыстық техниканы дауыспен басқару бағдарламалары *VoiceNavigator* мен *Truffaldino*;
- ұялы телефонды дауыс арқылы басқару кешені *DIVo*;
- ұзақтығы 3-5 секундты құрайтын парольді фразалармен тұлғаны сәйкестендіруге арналған *Voice Key* бағдарламалық модулі;
- ағылшын тіліндегі мәтінді жаздырту (диктовка) бағдарламалары: *Voice Type Dictation* (IBM), *DragonDictate* («*Dragon Systems*»); орыс тілінде *Комбат* («Вайт Групп») және *Диктограф* («*Voice Member Technology*»);
- *Microsoft Office XP-ге* енгізілген сөйлеуді тану жүйесі (тек ағылшын тілімен жұмыс істейді);
- дауыс арқылы іздеу (мысалы, *Google* іздеу жүйесі).

*Voice Navigator* бағдарламасы клавиатура көмегінсіз-ақ дауыс арқылы компьютерлік мәзірлерді басқаруға мүмкіндік береді. Бағдарламаны қолданар алдында микрофонға бұйыру мәзірлерін айту арқылы оны дайындап алу қажет (бұйыру мәзірлері кез келген тілде кез келген дауыспен қабылдана береді). Бағдарлама дауыс мәзірлерін танып бастау үшін тірек сөз бен сөз тіркестері арқылы оны «ояту» керек [8].

Сөйлеуді тану модульдерін қолдану іскерлік көп саласында: тұтынушыларға қызмет көрсетуде, сот сараптамаларын жүргізуде, биометрия, оқыту мен зерттеу жұмыстарында тиімді. Сөйлеу технологиясын жаппай енгізуді оған өте көп қаражаттың керектігі және әзірге оның сапасының төмендігі тежейді.

Жалпы алғанда, сөйлеуді автоматты талдау мәселесі өте күрделі және бүгінгі күні оның кейбір жақтары ғана шешілген. Оған қарағанда, сөйлеуді автоматты синтездеу мәселесі қарапайымдау және оның тұрмыста қолданылуын күнделікті кездестіруге болады. Келесі жағдайларда сөйлеу автоматты синтезделеді:

- телефон арқылы ағымдағы уақытты айту;
- метродағы аядамаларды хабарлау;
- мобильді операторлардың шоттағы қалған ақша сомасын айту және басқа да қызметтері;
- азаматтық қауіпсіздік жүйелерімен таныстыру және т.б.

Сөйлеудің автоматты синтезі бүгінгі күні алдын ала диктормен жазылған жекелеген дыбыстардан сөз және фразаларды құру (ком-

пилятивті синтез әдісі) және адамның сөйлеу трактын модельдеу (формантты-дауыстық әдіс) жолдары арқылы жүзеге асады [8].

Алғашқы аталған әдіс аз көлемді және алдын ала белгілі болған фразаларды синтездеуге арналған. Синтезделетін сөйлеу адамның табиғи сөйлеу тілінің элементтеріне негізделгендіктен дыбысталу сапасы өте жоғары болады. Дегенмен дыбыстарды біріктіру кезінде естірліктей кемшіліктер байқалады. Оған қоса түрлі интонациямен айтылатын фонемалардың ерекшеліктерін ескере отырып, дыбыс фрагменттерінің базасын құрастыру күрделі еңбекті талап етеді.

Екінші әдіс күрделі болып есептеледі, өйткені мұнда адамның сөйлеу тактысын нақты модельдеу және сөйлеудің интонациялық модуляциясын ескеру қажет. Аталған ерекшеліктеріне орай формантты-дыбыстық үлгіде синтезделетін сөйлеу дыбыстарының нақтылық деңгейі төмен.

Сөйлеудің синтезделуіне мысал ретінде Govorilka (құрастырушы: А. Рязанов, бағдарламаның тегін нұсқасы көрсетілген мекен-жайда енгізілген <http://www.vector-ski.com/vecs/govorilka>). Бағдарламаның басты ерекшеліктері мынадай:

- мәтін әртүрлі дауыспен әртүрлі тілдерде оқылады, орыс тілінде де;
- оқуға арналған түпкі мәтін мәтіндік файлға жүктелген немесе клавиатура көмегімен бағдарлама терезесінде енгізілген болуы мүмкін;
- сөйлеу синтезінің нәтижелерін WAV немесе MP3 форматындағы файлды басып алу арқылы сақтауға болады.

Бүгінгі компьютерлердің мүмкіндігінің жоғарылығына қарамастан, компьютерлерді толыққанды интерфейспен қамтамасыз ету әлі мүмкін болмай тұр.

### Семинар тапсырмалары

1. Таңба дегеніміз не? Тілдік және сөйлеу таңбалары арасындағы айырмашылықтарды қалай ажыратуға болады?
2. Қай салада дыбыстық сөйлеуді сөзбе-сөз танудың шектеулеріне назар аудармауға болады? Бұл шектеулер қай салалар үшін аса маңызды болып саналады?
3. Дыбысқа тән қандай артикуляциялық ерекшеліктерді білесіз?

4. Дыбыстық сөйлеуді талдаудың күрделі жүйесіне қай сала мамандары тартылуы қажет?
5. Сөйлеуді автоматты талдау құралдары қолданылатын қандай бағдарламаларды білесіз? Олардың қажеттілігі, тиімділігі және болашағы неде?
6. Дыбыстық сөйлеуді дәл анықтаудың құралдары және жолдары қандай?
7. Мәтінді автоматты талдау үдерісінің тарихын анықтаңыз.
8. Спектрограмма арқылы өз мысалдарыңызды талдаңыз.

### Әдебиеттер:

#### *Негізгі:*

1. t Hart J., Collier R. & Cohen A. A perceptual study of intonation: an experimental-phonetic approach to speech melody. – Cambridge: Cambridge University Press, 1990. – 212 p.
2. Кейспер К. О семантических основаниях описания Е.А. Брызгуновой русской интонации. – В кн.: Проблемы фонетики II. – М.: Наука, 1995. – С. 216-227.
3. Оде С. Что такое тональный акцент? // Жизнь языка и язык в жизни / Сб. статей, посвященный юбилею д.ф.н., проф. Э.Д. Сулейменовой: сост. Н.Ж. Шаймерденова, З.К. Сабитова, О.Б. Алтынбекова, М.И. Акберди; Под общ. ред. Н.Ж. Шаймерденовой. – Алматы: Қазақ университеті, 2005. – С. 76-85.
4. Оде С. Экспериментально-фонетический анализ тонов в папуаском языке Мпур // Материалы Международной конференции. 100 лет Экспериментальной фонетике в России. – Санкт-Петербург, 2001. – С. 137-140.
5. Оде С. Перцептивный анализ интонации. – В кн.: Проблемы фонетики. Сб. статей / Отв. ред. Т.М. Николаева. – М.: Прометей, 1993. – ч. 1. – 318 с. – С. 199-211.
6. Карагойшиева Д.А. Экспериментально-фонетический анализ интонации казахского языка. Алматы: Сору Land, 2008. – 232 с.
7. Марчук Ю.Н. Компьютерная лингвистика: учеб. пособие. – М.: АСТ Восток - Запад, 2007.
8. Потапова Р.К. Речевое управление роботом: лингвистика и современные автоматизированные системы. – М., 2005.
9. Фролов А.В., Фролов Г.В. Синтез и распознавание речи. Современные решения: электронный учебник. – <http://frolov-lib.ru/books/hi/ch00.html> (қаралған күні: 28.02.2012).

## **5-дәріс. Мәтінді автоматты түрде тану**

**Қарастырылатын мәселелер:** *жазба мәтінді автоматты талдау; дыбыстық тілді талдау; жазба мәтінді компьютерге енгізу; Оcr-бағдарламалар көмегімен мәтінді тану; мәтіннің келесі ерекшеліктері.*

**Тірек сөздер мен сөз тіркестері:** *жазба мәтінді автоматты талдау, дыбыстық тілді талдау, жазба мәтіндерді автоматты өңдеу, сканер, мәтіндік редактор, жасанды интеллект.*

Жазба мәтінді автоматты талдау, дыбыстық тілді талдау секілді оның компьютерге енгізілуінен басталады. Қазіргі кезде адамдар көбінесе жазба тасымалдағыштағы ақпаратпен кездеседі, сондықтан еңбекте жазба мәтіндерді автоматты өңдеу мәселесіне бірнеше параграфтар арналады. Бірінші параграфта мәтінді компьютерге енгізу және онымен байланысты жазба мәтінді тану үдерісіне кеңінен тоқталамыз.

Ақпаратты компьютерге енгізуде арнайы құрылғылар қолданылады – пернетақта, тышқан және т.б., ең қолайлы жолы көлемді мәтіндерді сканер арқылы енгізу болып саналады.

**Сканер** – фотоаппарат қағидасы бойынша жұмыс істейтін құрылғы, себебі ол мәтінді фотосурет түрінде көреді [1, 53-б.]. Мұндай мәтіндерді компьютер «түсіну», яғни графикалық күйдегі, әрқайсысының жеке екілік коды бар символдарды (мысалы, ASCII кодтау жүйесінде) мәтіндік қалыпқа келтіру үшін мәтінді автоматты тану бағдарламалары қажет (OCR = Optical Character Recognition).

Кез келген тыныс белгі және таңбалар символ болып есептеледі (апострф, тырнақша, тире, жақша). Сөз екі көршілес аралық (пробел) арасындағы символдар тізбегі ретінде түсіндіріледі.

Мәтінді автоматты тану бағдарламасы (OCR-бағдарлама) – болашақта мәтінді кез келген мәтіндік редакторда өңдеуге болатындай қағаз күйінен электронды күйге айналдыратын компьютерлік бағдарлама [1, 53-б.]. Бұндай бағдарламалар сканермен қоса жүреді, бірақ ең көп қызметке ие FineReader (Abbyy компаниясы) мен CuneiForm (Cognitive Technologies фирмасы) OCR-бағдарламалары.

Мәтінді автоматты танудың басқа да бағдарламаларымен мына веб-сайт мекенжайы арқылы таныса аласыз <http://kompkimi.ru/?p=617> (мерзімі: 02.02.2012).

OCR-бағдарламалары арқылы мәтінді танудың нәтижелері барынша нақты, бірақ мәтінді танудағы кейбір қиындықтар қателіктерге алып келеді, олар қолмен дұрыстауды қажет етеді.

Мәтінді танудың жазба мәтіннің келесі ерекшеліктерімен байланысты болуы мүмкін:

- әртүрлі шрифтердің қолданылуы;
- мәтінде бірнеше тілдің қолданылуы;
- мәтіндердің бірнеше бағанада орналасуы;
- мәтіннің ішінде кестелер мен суреттердің болуы;
- белгілердің ауытқуы (ажырауы, не болмаса бірігіп кетуі);
- кескінге тыс заттардың кіруі [1, 53 – 54-б.; 1, 34-б.].

Аталған қиындықтарды шешу үшін мәтінді тану бағдарламасы мәтіннің мағынасын тани алуы қажет, яғни OCR-бағдарламаның жасанды интеллект дәрежесін көтеру керек.

Жалпы алғанда, сапасы орташа және жоғары мәтіндерді тануда OCR-бағдарламалардың мәтінді тануы 99 %-ды құрайды, сондықтан бұл мәселе толыққанды шешілген деп пайымдауға негіз бар.

### **Семинар тапсырмалары**

1. OCR-бағдарламаларының басты мүмкіншіліктерін атаңыз.
2. OCR-бағдарламаларының даму мүмкіндігі қаншалықты?
3. Мәтінді автоматты анықтау жүйелерінің біреуінің артықшылығын атаңыз.
4. Мәтінді автоматты түрде талдаудың қиыншылықтары неде? Қалай ойлайсыз, оларды жеңілдетуге немесе жоюға бола ма? Жолдарын тауып, дәлелденіз.

### **Әдебиеттер:**

*Негізгі:*

1. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
2. Башмаков И.А., Башмаков А.И. Интеллектуальные информационные системы. – М., 2005.

*Қосымша:*

1. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд., испр. – М., 2007. – С. 47.

2. Сканирование и распознавание: профессиональный подход: материалы курса дистанционного обучения. – [www.online-academy.ru/scan.htm](http://www.online-academy.ru/scan.htm) (қаралған күн: 02.02.2012).

## **6-дәріс. Мәтінді автоматты түрде аннотациялау және рефераттау**

**Қарастырылатын мәселелер:** *мәтінді автоматты түрде аннотациялау және рефераттау ұғымы; реферат түрлері; автоматты аннотациялау жүйелері; аннотациялау және рефераттаудың келесі әдістері.*

**Тірек сөздер мен сөз тіркестері:** *мәтін, автоматты түрде аннотациялау және рефераттау, реферат, аннотация, аннотациялау жүйелері, құжат, байланысқан мәтін, реферат-клише, квазиреферат, мәтін компрессиясы, статистикалық әдіс, позициялық әдіс, логика-семантикалық әдіс.*

Адамды қоршаған әлемдегі мәтіндер санының өсіп келе жатқандығына байланысты мынадай мәселе туындайды: бұл ақпарат айдынында қалайша қажетті құжаттарды тауып және олардың мазмұндарымен танысуға болады? Берілген мәселені шешуде толық мәтінді құжаттардың рефераттары мен аннотацияларын құру қол ұшын бере алады. Олар оқырманға ағымдағы құжаттардың мазмұндары жайлы түсінікті ұсынады және әрбір жұмыстың толық мәтіндеріне назарын салу қажеттілігі дәрежесін бағалауға мүмкіндік жасайды. Одан бөлек рефераттар мен аннотациялар оқырман назарын жаңа мәліметтерге айрықша көңіл аудартады, яғни аз уақытта көптеген жаңа ақпаратты білуге мүмкіндік береді.

Рефераттар мен аннотациялар ағымдағы мәтін авторының немесе библиографиялық қызметшінің өз қолымен, сондай-ақ автоматты түрде арнайы компьютерлік бағдарламалардың көмегімен құрастырылады. Рефераттар мен аннотациялардың алғашқы түрлері мейлінше сапалы болып табылады. Себебі тек осындай жағдайда айтылымның негізгі ойын атайтын және байланысты сипаттамамен ерекшеленетін жазба мәтін құрылады. Бірақ аз уақыттың ішінде мәтіннің үлкен бөлігін өңдеу үшін мәтінді ре-

фераттау мен аннотациялау мәселелерін шешетін автоматтық құралдарды тарту қажеттігі туады.

Реферат байланысқан мәтін ретінде:

– *орталық тақырып*;

– *зат*;

– *мақсат*;

– *әдіс*;

– *зерттеу нәтижелерінің қысқаша мазмұнын анықтайды* [1, 55].

Рефераттарды, әдетте ғылыми-техникалық құжаттарға, яғни ғылыми монографиялар, мақалалар, ойлап шығарылған заттарға патенттер және т.б. құрастырады.

Ағымдағы мәтіннің жанры (монография, мақала, патент, т.б.) мен пәндік аясына қатысты (медицина, химия, лингвистика, т.б.) рефераттың сұралған элементтері өзгерген болуы мүмкін. Осылайша, ғылыми рефераттарға рефераттың жоғарыда аталған элементтеріне қосымша мән, тәжірибелік апробация, зерттеудің болашағының қысқаша мазмұны қосылады [2, 93 – 94-б.].

Рефераттың түрлері былайша түрленеді [2, 89-б.]:

– ***байланысқан мәтін*** – ағымдағы мәтіннің логика-мағыналық анализі негізінде туындайтын жаңа мәтіндік құрылым;

– ***реферат-клише*** – келтірілген клише қолданылған құрылымның модификациясы, бос ұяшықтар, оларда берілген мәтіннің анализінен кейін толтырылады;

– ***квазиреферат*** – мәтіннің мейлінше ақпаратты сөйлемдерінің тізімі.

Көрініп тұрғандай, рефератты байланысқан мәтіндерді автоматты құруға реферат-клише мен квазирефератты құруға қарағанда күрделірек компьютерлік бағдарламалар талап етіледі.

Кейбір зерттеушілер реферат пен аннотацияны синоним деп есептейді [2, 88-б.], ал кейбірі аннотацияны оның тақырыбы жайлы жалпы көрінісін байқататын құжат мазмұнының қысқаша мазмұны ретінде түсініп, екеуін ажыратуды ұсынады [1, 55-б.]. Осы анықтамаға сәйкес, оқырманды құжат мазмұнының желісімен таныстыратын рефератқа қарағанда, аннотация тек сигналды қызмет атқарады (белгілі тақырыпқа басылым бар).

Мәтіннің қысқа мазмұнын автоматты құрастыруға бағытталған көпшілік бағдарламаларда мәтін компрессиясының (ығыстырудың) түрлі деңгейлерін білдіруге болады, яғни бір ғана бағдар-

лама қысқаша рефераттарды да, қысқа аннотацияларды да жасайды. Осыған байланысты мәтіннің қысқа мазмұнын автоматты құру үдерісінде екі ұғым қолданылады: *мәтінді автоматты түрде аннотациялау және рефераттау*.

Рефераттау және аннотациялау үдерісінде құрастырылатын рефераттар мен аннотациялар екінші деңгейдегі құжаттарды білдіреді. Алғашқы деңгейдегі (ағымдағы) құжаттар – кітаптардың өзі, мақалалар, патенттер, т.б.

Автоматты түрде аннотациялау және рефераттау бағдарламалары оны адамның қалай жасауына байланысты болып табылады. Адам үшін бұл үдеріс келесі кезеңдерді қамтиды:

– **дайындық**: мәтін тақырыбы, оның түсінігін анықтау;

– **аналитикалық**: мәтінді үзінділерге бөлу (абзац, т.б.) және әрбір үзіндіде болашақ рефераттың жоспарын құрайтын басты мағыналы сөздерді бөліп қарастыру;

– **реферат пен аннотация құрастыру**: байланысқан мәтінге бөлініп қарастырылған мағыналы бірліктерді қосу.

Ағымдағы мәтіннің басты мағыналық бірліктері *тірек сөздер*, *тірек сөз тіркестері*, *тірек сөйлемдер* болады. *Тірек сөз* – мәтіннің негізгі мазмұнына қатысты және онда бірнеше рет қайталанатын мәнді сөз. *Тірек сөз тіркесі* – бір немесе бірнеше тірек сөздердің тіркесімі. *Тірек сөйлем* – бірнеше (екі және одан көп) тірек сөздерден тұратын сөйлем [1, 57].

Ағымдағы мәтіннен тірек сөз тіркестері мен сөйлемдерді бөлу тәсілдеріне орай, мәтінді автоматты түрде аннотациялау және рефераттаудың келесі әдістері ажыратылады [1, 59]:

- *статистикалық*;
- *позициялық*;
- *логика-семантикалық*.

*Статистикалық әдісте* сөздің тірек категориясына тәндігі оның статистикалық сипаттамаларымен анықталады: тірек сөз, осы әдіске сәйкес мәтіннің бастапқы сөздері арасында мейлінше көп кездеседі. Тірек сөйлем, сәйкесінше, бір-бірінен аздаған қашықтықта орналасқан бірнеше тірек сөздерді қамтиды.

*Позициялық әдісте* сөйлемнің тірек болу принципі оның мәтіндегі орнына сүйенеді: тірек сөйлемдер тақырып, тақырыпшаға кіреді, мәтіннің басы мен аяғында тұрады.



Мәтіннің құрылысы мен семантикасы ескерілетін *логика-семантикалық әдістің* мақсаты сөйлемдердің қызметімен айқындалады. Осындай сөйлемдер ерекше түрде басқа сөйлемдермен байланысқан, сөйлемнің белгілі синтаксистік типін иеленетін, т.б. семантикалық мәнді сөздерді қамтиды.

Автоматты түрде аннотациялау және рефераттаудың мейлінше қарапайым жүйесі болып AutoSummarizev MS Word функциясы, Intelligent Text Miner, Oracle Contexti Inxight Summarizer (AltaVista іздеу механизмінің компоненті) (IBM) жүйелері есептеледі. Шынында, бұл бағдарламалардың мүмкіндіктері ағымдағы құжаттың соны үзінділерін таңдаумен және олардың қысқа мәтіндерге қосылуымен шектелген [3].

Одан бөлек, мәтінді автоматты түрде аннотациялау және рефераттау жүйелерінің мысалдарын келтіруге болады:

– ОРФО 5.0 («Информатик» компаниясы): бағдарлама орыс мәтіндерін автоматты аннотациялау қызметін атқарады;

– «Либретто» («Медиалингва» компаниясы): бағдарлама Wordқа сәйкес құрылады және орыс, ағылшын мәтіндерін автоматты түрде аннотациялау және рефераттауды қамтамасыз етеді;

– «Следопыт» іздеу жүйесі құжаттарды автоматты түрде аннотациялау және рефераттау құралдарын қамтиды;

– Extractor мен TextAnalyst («Медиасистемы»компаниясы) синтаксистік анализаторлар көмегімен бөлінген атаулы топтар тізбегін білдіреді.

Жалпы, автоматты рефераттар мен аннотациялар квазирефераттар болып табылады, яғни мәтіннің автоматты компрессиясының нәтижесі көп жағдайда, тірек сөздердің жинағы, не тірек сөйлемдердің тізіміне айналады. Солай бола тұрса да квазирефераттар, аз уақыт ішінде үлкен көлемді аннотациялау және рефераттау мәселесін шешуге айтарлықтай дәрежеде көмектеседі.

### **Семинар тапсырмалары**

1. Мәтін рефератын құрастыру кезеңдерін сипаттаңыз.
2. Мәтінді автоматты түрде аннотациялау және рефераттаудың белгілі жүйелерін сипаттаңыз.
3. Мәтінді автоматты түрде аннотациялау және рефераттау жүйелеріне қандай мәселелер тиімді болып табылады? Жауабыңызды дәлелдеңіз.

4. Әртүрлі жанрдағы 2-3 мәтінге автоматты түрде аннотациялар мен рефераттарды жасап келтіріңіз.

### Әдебиеттер

#### Негізгі:

1. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб.пособие. – М.: Академия, 2004.
2. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. Пособие. – СПб.: Книжный Дом, 2007.
3. Хан У., Мани И. Системы автоматического реферирования // Открытые системы. 2000. – №12. – [www.osp.ru/os/2000/12/178370](http://www.osp.ru/os/2000/12/178370) (қаралған күн: 02.02.2012).

#### Қосымша:

1. Башмаков И.А, Башмаков А.И. Интеллектуальные информационные системы. – М., 2005. – С. 77-90.

## 7-дәріс. Мәтіннің автоматты анализі және синтезі

**Қарастырылатын мәселелер:** *мәтіннің графемалық, морфологиялық, синтаксистік, семантикалық анализдері; токенизация, парсер ұғымдары; формальді грамматика; машиналық негіз, машиналық жалғау; мәтіннің автоматты анализі.*

**Тірек сөздер мен сөз тіркестері:** *мәтіннің автоматтық анализі, токенизация, парсер, формальді грамматика; машиналық негіз, машиналық жалғау, стемминг, морфологиялық анализ, синтаксистік анализ, семантикалық анализ, семантикалық синтез.*

Мәтіннің автоматтық анализі берілген алгоритмге сәйкес, табиғи адамзат тілінде мәтінде компьютер орындайтын күрделі үдерістердің қатарын қамтиды. *Автоматтық анализде* мәтін компьютерге түсінікті оның лексема-морфологиялық, синтаксистік пен семантикалық түрлеріне өзгереді [1, 14-б.]. Табиғи тілде мәтінге лексема-морфологиялық, синтаксистік пен семантикалық компьютерлік көріністердің кері айналу үдерісі ***мәтіннің автоматтық синтезі*** деп аталады.

Мәтіннің автоматтық анализі мен синтезі компьютерлік лингвистиканың теориялық даму (жасанды интеллект жасаудың лингвистикалық негіздерін құрастыру) және адамның тәжірибелік қа-

жеттіліктерін іске асыру тұрғысынан да маңызды мәселесіне айналады, мысалы, машиналық аударманың тиімді жүйелерін құру.

Мәтіннің автоматтық анализі бірнеше кезеңдерді қамтиды:

– *графемалық анализ*: сөздер, сөйлемдер, абзацтар шекараларының және мәтіннің басқа да элементтерінің ажыратылуы (мысалы, газет мәтінінің қимасы);

– *морфологиялық анализ*: мәтінде әрбір қолданысқа түскен сөздің бастапқы тұлғасын анықтау және осы сөздің морфологиялық сипаттарының жинағы;

– *синтаксистік анализ*: мәтін сөйлемінің грамматикалық құрылымын айқындау;

– *семантикалық анализ*: фразалар мағыналарын анықтау [2, 94 – 106 –107-б.].

*Графемалық анализ токенизация* ретінде анықталады (ағылш. token = жеке сөз, фраза немесе мәтіннің өзге де маңызды элементі). Мәтіндік элементтердің шекараларының формальді сигналдары түрлі тектегі ажыратылымдар болады: сөздер арасындағы шекараны білдіретін бос орындар, сөйлемдер мен сөйлемдердің құрылымды бөліктері арасындағы шекараларды білдіретін *бас әріптер мен тыныс белгілері*, мағына арқылы байланысқан сөйлемдер топтарының, т.б. Араларындағы шекараларды білдіретін *абзацтық кері шегінулер* [3, 79-б.].

Дегенмен сөздердің шекараларын анықтаудың формальдық әдісі жиі қолданыла бермейді. Мәселен, қытай тілінде сөздердің формальдық шекаралары болмайды [4]. Бұдан бөлек, тіпті, таралымы кең еуропа тілдерінде де бос орынмен бөлінген, мысалы, New York тәрізді бір лексема түрінде түсінілетін сөздердің тұрақты тіркестері бар. Мысалы, көпсөзді лексема тізімдерін құру жолында осындай жағдайларды графемалық анализ жүйесінде есепке алу жөн.

*Морфологиялық анализде* мәтінде қолданысқа түскен әрбір сөз алғашқы тұлғасына келтіріледі және сөздің мәтіндік тұлғасының морфологиялық мінездемелерінің жинағы анықталады: сөз табы, зат есімдер үшін тегі (род), жекеше/көпше түрлері, септелуі, етістік үшін жекеше/көпше түрлері, жақ, т.б.

Мәтінде қолданылған әрбір сөз *сөз тұлғасы* (немесе *сөз қолданысы*) деп аталады. Мәтін байланыстылығын қамтамасыз ету үшін сол сөздердің қайталануы талап етіледі, сондықтан мәтіннің

бір немесе бірнеше сөйлеміндегі әртүрлі сөз тұлғалары бір ғана алғашқы тұлғаға келтіріледі. Мысалы:

Құлақтан кіріп бойды алар,  
Әсем ән мен тәтті күй.  
Көңілге түрлі ой салар,  
Әнді сүйсең менше сүй

(Абай).

Өлеңнің осы үзіндісіндегі сөз тұлғаларының әліпбилік-жиілік сөздігі былайша көрініс тапқан: *алар* – 1, *ән* – 1, *әнді* – 1, *әсем* – 1, *бойды* – 1, *көңілге* – 1, *кіріп* – 1, *күй* – 1, *құлақтан* – 1, *мен* – 1, *менше* – 1, *ой* – 1, *салар* – 1, *сүй* – 1, *сүйсең* – 1, *тәтті* – 1, *түрлі* – 1. *Ән* мен *әнді* сөз тұлғаларының түбірі бір, *ән* – атау септігінде, *ән+ді* – табыс септігінде тұр. *Мен* және *менше* сөз тұлғаларының түбірі – *мен*. *Мен* – жалғаулық шылау, ал *мен+ше*дегі *мен* – жіктеу есімдігі, *-ше* зат есімнен үстеу тудырушы жұрнақ. *Сүй* мен *сүйсең* сөз тұлғаларының түбірі бір. *Сүй* – етістік, *сүй+се+ң* сөз тұлғасындағы *сүй* – етістік, *-се* – шарты рай формасы, *-ң* жіктік жалғаудың II жағы.

Бізге таныс сөздіктерде, әдетте сөз тұлғалары емес, белгілі алғашқы тұлғаға келтірілген сөздер санамаланады. Мәтінде пайдаланылған сөз тұлғаларының осындай алғашқы тұлғасы ретінде, тіл типіне қатысты, *лемма* (лексеманың сөздіктегі тұлғасы) немесе негіз (сөз түрлендіретін морфемасыз сөздің негізгі бөлігі) қызмет атқара алады. Мысалы, *swim*, *swims*, *swam* мен *swimming* ағылшын сөз тұлғалары *swim* деген бір ғана леммадан шыққан.

**Машиналық жалғау** – нақты лексеманың форма түрленуін сипаттайтын және парадигмалар түрінде ұсынылатын элементтер. Машиналық жалғаудың барлық мүмкін жинақтары лексеманың *типтік парадигмасында* анықталған. Сонда да, бір жағынан, әртүрлі лексемалардың типтік парадигмаларының (машиналық жалғау да) сәйкес келетінін байқауға болады [7, 144–145-б.].

Көпмағыналылықты автоматты шешу немесе талдаудың бірнеше нұсқаларына жол беретін сөз тұлғасының дұрыс интерпретациясын таңдау деп түсінілетін *омонимияны шешу* қолмен жасау немесе болжалды модельдер негізінде автоматты (мәселен, ағылшын тілінде белгісіздік артиклі мен зат есімнің одан кейінгі-

мен тіркесімі) немесе адам жасаған ережелер негізінде орындалады. Осындай ережелердің мысалдары:

– егер сөз тұлғасы етістік те, зат есім де бола алса және оның алдынан артикль тұрса, бұл сөз тұлғасы берілген жағдайда зат есім болады;

– егер сөз тұлғасы септеулік жалғау да, шылау да болса және одан кейін сөйлем соңына дейін етістік кездеспесе, бұл сөз тұлғасы берілген жағдайда септеулік жалғау (предлог) болады [6].

Автоматты морфологиялық анализге *парсерлер* – сөздердің автоматты анализіне арналған арнайы компьютерлік бағдарламалар қолданылады [8]. Морфологиялықтан басқа сөйлемдердің синтаксистік құрылымын анықтауда автоматты талдауда синтаксистік парсерлер бар.

Жалпы, морфологиялық анализ келесі кезеңдерден тұрады [6]:

– лемматизация түрін иеленетін сөз тұлғаларын қалыптандыру, яғни кейбір ортақ түсінікке – алғашқы тұлға (лемма) немесе стеммингке әр алуан сөз тұлғаларының дерегі, яғни бір квазинегізге түрлі сөз тұлғаларының келтірілуі;

– тіл бөлігінің тэгингі, яғни сөз тұлғаларының мәтіндегі кез келген сөз тұлғаларына нұсқау;

– толық морфологиялық анализ – грамматикалық сипаттамаларды сөз тұлғаларына телу.

**Синтаксистік анализде** міндетті түрде сөйлемдегі сөздердің рөлін және өзара байланыстарын анықтау керек. Автоматтық анализдің бұл кезеңінің нәтижесі модель түрінде әрбір сөйлемнің синтаксистік байланыстарын, мысалы, тәуелділік бәйтерегі бейнесінде келтіру болып табылады.

**Семантикалық анализ** мәтіннің автоматтық анализінде күрделірек бағытты білдіреді. Осындай жағдайда, мәтіндегі сөздер арасындағы семантикалық қатынастарды орнату, бір ғана түсінікке қатысты түрлі тілдік ұғымдардың бірлігін, т.б. талап етеді.

Сөйлемдердің семантикалық анализі үшін септік грамматикалары мен семантикалық септіктер (валенттіліктер) қолданылады. Осындай жағдайда, сөйлем семантикасы басты сөздің (етістік) байланысының семантикалық актанттарымен сипатталады [10, 96-б.].

Семантикалық анализдің негізінде сөз негізі элементарлы синтаксистік бірлік болып табылмайды. Ол одан да ерекше элементарлы мағыналарға – семантикалық тілдің сөздігінің бірлікте-

ріне бөлінеді. Семантикалық тілдің бұл бірліктері «молекулалар» – табиғи тілдің реалды сөздерінің мәні жинастырылатын түрлі комбинациялардан тұратын өзіндік атомдары болады [9, 254-б.].

Дәл семантикалық анализ тілдің барлық деңгейлеріндегі автоматты анализде туындайтын көпмағыналылық (омонимия) мәселелерін шешуге мүмкіндік береді.

- Лексикалық омонимия: ортақ мағыналық элементтері жоқ жазылған сөздер немесе дыбысталу жағынан сәйкес келуі, мысалы, алма – жеміс және етістіктің болымсыз түрі.

- Морфологиялық омонимия: бір сөз (лексема) формаларының сәйкес келуі, мысалы, пол сөз тұлғасы пол зат есімінің атау мен табыс септігіне сәйкес келеді.

- Лексика-морфологиялық омонимия (омонимияның жиі кездесетін түрі): екі түрлі лексемалар сөз тұлғаларының сәйкес келуі.

- Синтаксистік омонимия: бірнеше интерпретациясы бар синтаксистік құрылыстың бір мағыналы еместігі, *Flying planes can be dangerous* (Хомскийдің танымал мысалы, онда *Flying* сөз тұлғасы сын есім немесе зат есім түрінде интерпретациялана алады) [10;11, 93-94].

Автоматты синтез байланысқан мәтіннің жасалу үдерісін білдіреді, оның жеке кезеңдері морфологиялық анализ сияқты болады. Алайда кері ретпен орындалады: алдымен, семантикалық синтез, кейін синтаксистік, морфологиялық пен графемалық іске асады.

**Семантикалық синтез** мағыналық жазбаның оның синтаксистік құрылысына өтуін білдіреді; синтаксистік – фразаның синтаксистік құрылысынан фразаны білдіретін сөз тұлғаларының лексика-грамматикалық сипаттама тізбегіне өтуі; *лексика-морфологиялық* – лексика-грамматикалық сипаттаманың реалды сөз тұлғасына өтуі [11]. Морфологиялық синтезде сөздің қалыпты формасы мен оның параметрлері бойынша, бағдарлама сәйкес сөз тұлғасын табады. Графемалық синтез сөзді ортақ мәтінге біріктіреді, ену мәтінінің үзіндісіне шығу мәтінінің үзіндісінің сәйкес келуін бақылайды [12, 108-б.].

Америка ғалымы Джозеф Вейценбаумнің «Элиза» бағдарламасы ағылшын тіліне жазба диалогын синтездейтін компьютерлік бағдарламалардың алғашқысы болып табылады. Оның бірінші нұсқасы 1966 жылы сынақтан өткен. Бұл бағдарлама активті

естілім техникасын жүзеге асыра отырып, психотерапевтпен диалогты имитациялаған. Бағдарлама Бернард Шоудың «Пигмалион» пьесасының бас кейіпкері «жоғарғы санатты адамдар» тіліне үйреткен Элиза Дулитлге байланысты қойылған. Бағдарлама мақсаты – дәл мағыналық ойлаудың моделін көрсетпеу. Шектеулі бағдарламалық ресурстар, сондай-ақ лингвистикалық анализ бен синтездің бастапқы деңгейімен түсіндірілетін тілдік әдепті модельдеу болған [13, 21-б.].

Бағдарлама лингвистикалық ақпараттың минимумын қосқан: 1) кейбір тұрақты тілдік формулаларын актуалдайтын тірек сөздердің кешені; 2) ортақ сұраққа алдыңғы айтылымды трансформациялау мүмкіндігі [сол жерден]. «Элиза» бағдарламасының жұмысының кескіні ретінде жас әйелдің сөйлесуі мен 1966 жылы Джозеф Вейценбаум жариялаған «Элиза» бағдарламасының келесі үзіндісі түсініледі. Бағдарлама арқылы модельденген айтылымдар бас әріппен ерекшеленген.

*Men are all alike.*

*IN WHAT WAY?*

*They're always bugging us about something or other.*

*CAN YOU THINK OF A SPECIFIC EXAMPLE?*

*Well, my boyfriend made me come here.*

*YOUR BOYFRIEND MADE YOU COME HERE*

*He says I'm depressed much of the time.*

*I AM SORRY TO HEAR YOU ARE DEPRESSED*

*It's true. I am unhappy.*

*DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY?*

*I need some help, that much seems certain [14 бойынша цит.].*

Бағдарлама жасалымында қолданылған алгоритмдердің қарапайымдылығына қарамастан, оның көмегімен практикалық түрде 1950 жылы ағылшын ғалымы Алан Тьюринг ұсынған белгілі тестін кері қайтаруға қол жеткізілді. *Тьюринг тестінің* болмысы келесі түрде формаланады: егер қарым-қатынас қатысушыларының көпшілігі кіммен әңгімелесіп отырғандығы – адаммен немесе машинамен екендігін анықтай алмаса, жасанды интеллект құру мәселесі шешілген болады. «Элиза» бағдарламасымен тәжірибе жүргізгенде кіммен әңгімелесіп отырғандығын анықтау ұсынылған. Нәтижесінде сыналушылардың 62 %-ы оларға адам деп жауап берген [13, 21].

Табиғи тілде жазба диалогтарын компьютерлік модельдеу жүйелерінің қазіргі кезеңінде күрделірек алгоритмдерді қолдануға енгізген. Жекелей алғанда, виртуалдық агенттерді (немесе боттар) жасауға қолданылған AIML (*Artificial Intelligence Markup Language*) жасанды интеллекті үшін арнайы белгілеу тілі құрылған. Өңгімелесушімен диалогты модельдейтін боттар пайдаланушының мобильді оператор мен сауда желісінің мүмкіндіктері жайлы сұрақтарының жауаптарын алу үшін компьютерлік ойындар мен корпоративті веб-парақшаларда қолданылады.

### Семинар тапсырмалары

1. Мәтіннің автоматтық анализі мен синтезі үшін релевантты табиғи тілдің деңгейлерін атап, қысқаша сипаттаңыз.
2. Мәтіннің автоматтық анализінің негізгі түсініктеріне анықтама беріңіз: *сөз, сөз формасы, лемма, машиналық негіз, стемминг, сөз табы тэгин, парсер, Тьюринг тесті*.
3. Мәтіннің автоматтық анализінің кезеңдерін атап, қысқаша сипаттаңыз.
4. Мәтіннің автоматтық синтезінің кезеңдерін атап, қысқаша сипаттаңыз.
5. Диалогтар, сондай-ақ авто-робот жауап берушілердің компьютерлік модельдеу жүйелерін сипаттаңыз. Роботтарды оқыту қалай жүзеге асады? Авто-робот жауап берушіні қалай танып білуге болады?

### Әдебиеттер:

#### Негізгі:

1. Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998.
2. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011.
3. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.
4. Huang, C., Simon, P., Hsieh, S., & Prevot, L. Rethinking Chinese Word Segmentation: Tokenization, Character Classification, or Word break Identification // Proceedings of the Association for Computational Linguistics. Demo and Poster Sessions. Prague, 2007. P. 69-72. <http://www.aclweb.org/anthology/P/P07/P07-2018.pdf> (қаралған күн: 28.02.2012).
5. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
6. Толдова С.Ю., Бонч-Осмоловская А.А. Автоматический морфологический анализ//Фонд знаний «Ломоносов». – М., 2011. [www.lomonosov-fund.ru/enc/ru/encyclopedia:0127430](http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127430) (қаралған күн: 28.02.2012).
4. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. – СПб.: Книжный Дом, 2007.



5. Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28-31 октября 2010 г.) / отв. ред. В.А. Баранов. – Уфа; Ижевск: Вагант, 2010. – С. 167-171.
6. Апресян Ю.Д. Идеи и методы современной структурной лингвистики. – М.: Просвещение, 1966.
7. Большакова Е.И. Компьютерная лингвистика: методы, ресурсы, приложения // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 90-105.
8. Мельчук И.А. Автоматический синтез // Большая советская энциклопедия. – М.: Советская энциклопедия, 1969-1978.
9. <http://dic.academic.ru/dic.nsf/bse/61319/> Автоматический (қаралған күн: 28.02.2012).
10. Клышинский Э.С. Начальные этапы анализа текста // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 106-140.
11. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
12. Maher J. Eliza // Digital Antiquaria, Interactive Fiction. June 15, 2011. <http://www.filfre.net/2011/06/eliza-part-1/> (қаралған күн: 28.02.2012).

#### Қосымша:

1. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд. – М., 2007. – С. 50-51, 66-67.

## 1-АРАЛЫҚ БАҚЫЛАУ БОЙЫНША ТЕСТ ТАПСЫРМАЛАРЫ

### 1. Әлеуметтік лингвистика терминдер сөздігіндегі тілге берілген анықтама?

А) тіл (ағыл. *language*) – ойлаумен, халықтың менталитетімен, мәдениетімен үздіксіз, тығыз байланыстағы адам қарым-қатынасының әлеуметтік аса маңызды құралы; шартты, бірақ құрылымдық негізделген мағынасы бар таңбалар мен нышандардың жүйесі. Тіл – адамдар арасындағы саналы әлеуметтік сөйлеу қарым-қатынасының негізі; адамдардың мінез-құлқын реттеудің, қоғамды және оның әралуан құрылымдық элементтерін басқарудың маңызды құралы;

В) тіл – өте күрделі, көп қырлы, сан салалы құбылыс;

С) тіл – адам баласы қоғамында қатынас, яғни сөйлесіп, пікір алысудың құралы ретінде пайда болды. Тарихта мұндай коллектив алғашында ру, тайпа, түрінде, кейінде халық тілі, ұлт болып өмір сүрді, осыған орай, ру тілі, тайпа тілі, халық тілі болды;

D) тіл – адамдардың ойлау қабілетін іске асырушы, оны одан әрі дамытушы, ерік-жігерін, сезімін білдіруші, мәдени-тарихи дәстүрлерін ұрпақтан-ұрпаққа жеткізуші құрал.

**2. Алғаш рет «тіл» (язык) мен «сөздің» (речь) айырмашылық-ерекшеліктерін, олардың байланысын атап көрсеткен ғалым?**

- A) Р. Мерингер
- B) Г. Шухардт
- C) Ф. де Соссюр
- D) К. Фосслер

**3.Тіл білімі нені зерттейді?**

A) тілдің қазіргі күйін, жағдайын, бұрынғы өткен тарихын;  
B) әлем тілдерінің құрылысын, элеуметтік функциясын және тарихи дамуын, адамзаттың табиғи тілін, оның жалпы заңдылықтарын, қолданылуын және адам қоғамын;

C) тілдің қоғамдық қызметі, стиль тармақтарының қолданылуы, тілдік нормалардың қалыптасуы мен өзгеруі, тілдің біркелкі қалыптасқан нормаларын қолдану шартын;

D) мәдени дамуы бір-біріне ортақ, бір-біріне ұқсамайтын тілдік жүйе құрылымын.

**4. Тіл білімі тілдің қоғамдық тәжірибесі қатысына қарай қандай түрлерге бөлінеді?**

- A) фундаменталды;
- B) фундаменталды, нормативті;
- C) нормативті, қолданбалы;
- D) фундаменталды, нормативті, қолданбалы.

**5. Тілдің қазіргі күйін, жағдайын, бұрынғы өткен тарихын зерттейтін тіл білімінің саласы?**

- A) нормативті;
- B) қолданбалы;
- C) фундаменталды;
- D) жалпы.

**6. Тіл білімінің бір-бірімен тығыз байланысты қанша саласы бар?**

- A) 2;
- B) 4;
- C) 3;
- D) 5.

**7. Мәдени дамуы бір-біріне ортақ, бір-біріне ұқсамайтын тілдік жүйе құрылымын зерттейтін тіл білімінің саласы?**

- A) сипаттамалы;
- B) қолданбалы;
- C) диахронды;
- D) жеке.

**8. Тіл білімі мен әдебиеттану ғылымы тоғысып, қандай жаңа саланың негізін салды?**

- A) әлеуметтік тіл білімі;
- B) стилистика;
- C) психолінгвистика;
- D) лингвомәдениеттану.

**9. Қолданбалы лингвистика термині қашан пайда болды?**

- A) XIX ғасырда;
- B) XX ғасырдың 40-жылдары;
- C) XX ғасырдың 30-жылдары;
- D) XX ғасырдың 20-жылдары.

**10. Табиғи тіл дегеніміз не?**

A) тарихи қалыптасқан, белгілі этникалық топ немесе ұлттық мемлекетте қолданылатын таңбалар жүйесі;

B) адамдар арасындағы саналы әлеуметтік сөйлеу қарым-қатынасының негізі;

C) адамдардың ойлау қабілетін іске асырушы, оны одан әрі дамытушы, ерік-жігерін, сезімін білдіруші, мәдени-тарихи дәстүрлерін ұрпақтан-ұрпаққа жеткізуші құрал;

D) адам баласы қоғамында қатынас құралы, сөйлесіп, пікір алысудың құралы.

**11. Айналадағы дүние туралы анықталмағанды (белгісізді) анықтайтын мәліметтер қалай аталады?**

- A) мәлімет;
- B) тәсіл;
- C) ақпарат;
- D) технология.

**12. Ақпарат берілуіне қарай қанша түрге бөлінеді?**

- A) 2;
- B) 3;
- C) 4;
- D) 5.

**13. Аудиалды ақпарат не арқылы беріледі?**

- A) дыбыс арқылы;
- B) сезім арқылы;
- C) иіс сезу арқылы;
- D) таңба арқылы.

**14. Адамдар арасында, тірі организмдерде, техникалық құрылғыларда және қоғамдық өмірде ақпаратты жеткізу, жинақтау және түрлендіру үдерістері қалай аталады?**

- A) технология;
- B) ақпараттық технология;
- C) ақпарат;
- D) ақпараттық үдерістер.

**15. Технология ұғымының қанша анықтамасы бар?**

- A) 180;
- B) 300-ге жуық;
- C) 250 шамасында;
- D) 325.

**16. Кибернетика ғылымының атасы Ноберт Винердің ақпаратқа берген анықтамасы қандай?**

- A) айналадағы дүние туралы анықталмағанды (белгісізді) анықтайтын мәліметтер;
- B) сыртқы дүниеден біздің оған бейімделуіміздің және біздің сезімдеріміздің бейімделуінің процесінде алынған мазмұнды көрсету;
- C) қоршаған ортадағы жүріс-тұрысын реттеу үшін адам, тірі ағзалар немесе техникалық жүйелер арқылы жеткізілетін қоршаған әлем туралы мәліметтер;
- D) кейбір зат немесе құбылысты білдіретін таңба.

**17. Кодтау дегеніміз не?**

- A) компьютер көмегімен ақпаратты алу;
- B) ақпаратты алу, сақтау, жеткізу, тарату;
- C) ақпаратты белгілі бір алфавит арқылы ұсыну;
- D) ақпаратты сақтау, қабылдау.

**18. Кодтау үдерісінде сөздер (символ) мен оның мағынасының өзара қарым-қатынасы қалай аталады?**

- A) семантика;
- B) грамматика;
- C) фонетика;
- D) семиотика.

**19. ЭЕМ көмегімен ақпараттың жинақталуы, өңделуі мен жеткізілуі жайлы ғылым қалай аталады?**

- A) физика;
- B) электроника;
- C) информатика;
- D) математика.

**20. Зерттеушілер ақпараттық төңкерістің қанша кезеңін ажыратады?**

- A) 5;
- B) 6;
- C) 2;
- D) 3.

**21. ЭЕМ және компьютер ұғымдарының айырмашылығы бар ма?**

- A) екеуі бір ұғым;
- B) бар;
- C) айырмашылықтар кездеседі;
- D) өте көп.

**22. Компьютер дегеніміз не?**

A) бағдарламалар мен олардың үзінділеріне, бағыныңқы бағдарламаларға, процедураларға, бағдарламалық модульдерге, машиналық бағдарламалар мен олардың үзінділеріне арналған жалпы атау;

B) орындалуға (өңделуге) тиіс реттелген командалар тізбегі, есеп шығару алгоритмін сипаттайтын бағдарламалу тілінің сөйлемдер жиыны

C) адам құрастырған алгоритм (немесе бағдарлама) бойынша машинаға түсінікті тілде жазылған ақпаратты автоматты құру, өңдеу, жеткізу, іске қосу үшін жұмыс істейтін электронды құрылғы;

D) ақпаратты енгізу/шығаруға арналған құрылғы.

**23. Соңғы қолданушыға арналған және ақпарат бойынша түрлі операцияларды орындауға мүмкіндік беретін бағдарлама қалай аталады?**

- A) жүйелік;
- B) қолданбалы;
- C) графикалық;
- D) функционалдық.

**24. Лингвистің АЖО-сы дегеніміз не?**

A) лингвист ақпаратының жартылай орындалуы;

B) лингвист адамның жүйеге оралуы;

- C) лингвистикалық ақпараттың жұмыс орны;
- D) лингвистің автоматты жұмыс орны.

**25. Мәтінді дигитализациялау дегеніміз не?**

- A) соңғы қолданушыға арналған және ақпарат бойынша түрлі операцияларды орындау;
- B) графикалық көріністерді өңдеу;
- C) статистикалық мәліметтерді өңдеуге арналған электронды кестелерді құру;
- D) жазбаша және ауызша формадағы тілдік материалды сандық қалыпқа түсіру.

**26. Мәтінді автоматты талдау үдерісі қанша кезеңнен тұрады?**

- A) 3;
- B) 5;
- C) 4;
- D) 2.

**27. Спектограмма дегеніміз не?**

- A) акустикалық сигнал;
- B) дыбысталудың ұзақтығы;
- C) дыбыстардың фото-бейнесі;
- D) дыбыстар жылдамдығы.

**29. Ұялы телефонды дауыс арқылы басқару бағдарламасы қалай аталады?**

- A) VoiceNavigator;
- B) DIVo;
- C) Voice Key;
- D) Truffaldino.

**29. Клавиатура көмегінсіз-ақ дауыс арқылы компьютерлік мәзірлерді басқаруға мүмкіндік беретін бағдарлама?**

- A) Voice Key;
- B) DIVo;
- C) Truffaldino;
- D) VoiceNavigator.

**30. Govorilka бағдарламасын құрастырған кім?**

- A) В. Гумбольдт;
- B) Г. Штейнталь;
- C) К. Фосслер;
- D) А. Рязанов.

**31. Болашақта мәтінді кез келген мәтіндік редакторда өңдеуге болатындай қағаз күйінен электронды күйге айналдыратын компьютерлік бағдарлама қалай аталады?**

- A) ASCII кодтау жүйесі;
- B) мәтінді автоматты тану бағдарламасы (OCR-бағдарлама);
- C) Voice Navigator бағдарламасы;
- D) Voice Type Dictation (IBM).

**32. Кез келген тыныс белгі және таңбалар символ бола ала ма?**

- A) жоқ;
- B) кейде;
- C) бола алады;
- D) бола алмайды.

**33. Фотоаппарат қағидасы бойынша жұмыс істейтін құрылғы қалай аталады?**

- A) пернетақта;
- B) тышқан;
- C) монитор;
- D) сканер.

**34. Мәтіннің мейлінше ақпаратты сөйлемдерінің тізімі қалай аталады?**

- A) байланысқан мәтін;
- B) реферат-клише;
- C) квазиреферат;
- D) аннотация.

**35. Автоматты түрде аннотациялау және рефераттау бағдарламаларында мәтін тақырыбы, оның түсінігін анықтау қай кезеңде қамтылады?**

- A) реферат пен аннотация құрастыру кезеңі;
- B) дайындық кезеңі;
- C) аналитикалық кезең;
- D) соңғы кезең.

**36. Мәтіннің негізгі мазмұнына қатысты және онда бірнеше рет қайталанатын мәнді сөз қалай аталады?**

- A) тірек сөйлем;
- B) тірек сөз тіркесі;
- C) тірек белгі;
- D) тірек сөз.

**37. Мәтінді автоматты түрде аннотациялау және рефераттаудың қандай түрлері бар?**

- A) статистикалық, позициялық;
- B) статистикалық;
- C) позициялық, логика-семантикалық;
- D) статистикалық, позициялық, логика-семантикалық.

**38. «Либретто» бағдарламасының атқаратын қызметі?**

- A) орыс мәтіндерін автоматты аннотациялау;
- B) құжаттарды автоматты түрде аннотациялау;
- C) орыс, ағылшын мәтіндерін автоматты түрде аннотациялау және рефераттау;
- D) синтаксистік анализаторлар көмегімен бөлінген атаулы топтар тізбегін құру.

**39. Синтаксистік анализ дегеніміз не?**

- A) мәтінде әрбір қолданысқа түскен сөздің бастапқы тұлғасын анықтау және осы сөздің морфологиялық сипаттарының жинағы;
- B) мәтін сөйлемінің грамматикалық құрылымын айқындау;
- C) фразалар мағыналарын анықтау;
- D) сөздер, сөйлемдер, абзацтар шекараларының және мәтіннің басқа да элементтерінің ажыратылуы.

**40. Машиналық жалғау дегеніміз не?**

- A) мәтінде әрбір қолданысқа түскен сөздің бастапқы тұлғасын анықтау және осы сөздің морфологиялық сипаттарының жинағы;
- B) табиғи тілде мәтінге лексема-морфологиялық, синтаксистік пен семантикалық компьютерлік көріністердің кері айналу үдерісі;
- C) машиналық аударманың тиімді жүйелерін құру;
- D) нақты лексеманың форма түрленуін сипаттайтын және парадигмалар түрінде ұсынылатын элементтер.



## КОМПЬЮТЕРЛІК ЛИНГВИСТИКА: ҚОЛДАНБАЛЫ САЛАЛАРЫ

### 8-9 дәріс. Машиналық аударма

**Қарастырылатын мәселелер:** *аударма және машиналық аударма түсінігі; машиналық аударма жүйелерінің классификациясы; машиналық аударма түрлері; машиналық аударма мәселелері;*

**Тірек сөздер мен сөз тіркестері:** *аударма, машиналық аударма, автоматты аударма, машиналық аударма жүйелерінің классификациясы, машиналық аударма түрлері, электрондық сөздік, онлайндық аудармашы, аудармашы сөздік.*

**Машиналық аударма** – бір табиғи тілдегі мәтінді басқа тілге арнайы компьютерлік бағдарламамен аудару процесі. Машиналық аудармаға арналған ғылыми зерттеу бағыттарының аттары да, әртүрлі жүйелерді құрастыру да машиналық аударма, машиналық аударма жүйелері деп аталады [1, 14-б.].

Машиналық аударма үдерістерін адамның қатысуына байланысты төмендегідей түрлерге бөлеміз:

– МАНТ (*Machine-Assisted Human Translation*) – аударманы бағдарлама (компьютер) құралдарын қолдану арқылы адамның өзі аударды (жүзеге асырады).

– НАМТ (*Human-Assisted Machine Translation*) – адамның қатысуы арқылы жүзеге асатын машиналық аударма.

– FAMT (*Fully-Automated Machine Translation*) – толық автоматтандырылған машиналық аударма.

Бірінші жағдайда аударманы жылдамдату мен жеңілдету мақсатында адам барлық компьютерлік құралдарды пайдаланады, алайда мәтіннің аудармасын адамның өзі жасайды. Көмекші құрал ретінде әртүрлі электрондық сөздіктер, терминологиялық деректер қоры келеді.

Машиналық аударма жүйесінің екінші түрі автоматтандырылған аудармаға жақын, мұнда аударманы машина жасайды, ал адам көмекші қызмет атқарады.

Машиналық аударманы жүзеге асыру үшін компьютерге аударма алгоритміне негізделген арнайы бағдарлама енгізіледі.

Аударма алгоритмі – ол бір мағыналық жүйе ретінде және аударма бағытында берілген тілдер жұбына (пара языков L1–L2) нақты бір тілден басқа тілге мәтін бойынша қатаң анықталған аударма сәйкестілігін табу үшін жүргізілетін іс-әрекет ретінде түсіндіріледі.

Машиналық аударма жүйелеріне баламалы, вариантты және трансформациялы аударма сәйкестіліктерін жеткізуді, беруді қамтамасыз ету үшін қажетті грамматикалық (морфологиялық, синтаксистік, семантикалық) ақпараттармен жабдықталған біртілді, көптілді, көптілді сөздіктер қосылады. Сонымен қатар машиналық аударма жүйесіне қандай да бір формалды грамматика мәтіндерін автоматты түрде өңдейтін, өңдеуді орындай алатын грамматикалық талдау құралдарының алгоритмдері де қосылады. Машиналық аударма жүйесінде талдау мен жинақтауды қамтамасыз ететін ең кең таралған формалды операциялар жүйелілігі, бірізділігі төмендегідей болып табылады:

1) 1-кезеңде компьютерге мәтін енгізіледі және компьютерге кіретін сөздіктен, яғни аударма жасалатын тіл сөздігінен сөз формалары (нақты грамматикалық формадағы сөзді, мысалы, жатыс септігіндегі, көпше түрдегі) морфологиялық талдаулармен бірге іздестіріледі, бұл кезде берілген сөз форманың белгілі бір лексемаға (сөз-сөздіктің бірлігі ретінде) жататындығы, тәуелділігі анықталады;

2) 2-кезеңде тұрақты сөз тіркестері, фразеологиялық бірліктер және берілген пән саласында қолданылатын штамп тілдік бірліктер аударылады (мысалы, ағылшын тілінен орыс тіліне аудару барысында *in case of*, *in accordance with* сияқты тілдік оралымдар

бірдей балама алады және бұрынғы грамматикалық талдау бойынша алынады. Компьютерге кіретін мәтін элементтерінің негізгі грамматикалық (морфологиялық, синтаксистік), семантикалық және лексикалық сипаттарын анықтау (мысалы, зат есімнің көпше және жекеше түрлерін, етістіктің шақтарын, сөз форманың берілген мәтіндегі синтаксистік қызметін және тағы басқа), одан кейін балама тізімі бойынша бір мағыналы сөздер аударылады, ал көп мағыналы сөздерді аудару үшін мәнмәтіндік деп аталатын сөздіктер және белгілі мәнмәтінге сұрақ қою алгоритмдері бар сөздік мағыналары қолданылады;

3) соңғы грамматикалық талдау, мұнда берілген компьютерден аударылып шығатын тілдің қатысуымен, қажетті грамматикалық ақпараттар анықталады (мысалы, орыс тіліндегі сани – шана, ножницы – қайшы сияқты зат есімдермен қатар тұрған етістік көпше түрде тұруы қажет, ал түпнұсқада ол жекеше түрде болуы мүмкін);

4) компьютерден аударылып шығатын сөз формалар мен тұтас сөйлемдерді компьютерден, машинадан шығатын тілде жинақтау. Нақты тілдік жұптардың морфологиялық, синтаксистік және семантикалық ерекшеліктеріне байланысты жалпы аударма алгоритмі аударма бағытында басқа да кезеңдерді басынан өткереді. Талдау мен жинақтау жеке сөйлемдерге немесе компьютер жадында сақталған бүкіл мәтінге жүргізілуі мүмкін; соңғы жағдайда аударма алгоритмі анафоралық (қайталанатын дыбыстар, ритмдер, сөздер) байланысты ескере отырып анықталады (мысалы, зат есімнің орнын басатын есімдікпен байланысты).

Адамның қатысуы арқылы жүзеге асатын машиналық аударма жүйелері: Prompt (бірнеше нұсқасы бар), L-MASTER 98, Magic Translator 3.0, Translate Now! 1.6 (beta), Instant Messenger Translator Ver.1.0, Translation Memory, Socrat, Ampar, Nerpa, Systran, Frap және тағы басқа электрондық сөздіктер бар. Бұл жүйелерде аударманы машина жасайды, ал адам кейбір қиын мәселелерді шешу үшін қатысады. Мұндай аударма жүйелерінде кәсіби аудармашылардың деректер қоры сақталып тұрып, келесі рет аудару үдерісінде дайын аударма берумен шектеледі.

**PROMT компаниясы** – қолданбалы лингвистика саласынан және еуропа тілдерінен аударма жасайтын автоматтандырылған технология жасаудан әлемдегі бірінші ұйымдардың бірі.

PROMT компаниясының мақсаты – әртүрлі тілдерде еркін қарым-қатынас жасайтын бағдарламалық (программалық) құрылғыларын құру. Бұл сөздіктің бірнеше нұсқалары бар.

*PROMT Translation Office 2000/ XT* жүйесін “Проект МТ” компаниясы құрастырған (Санкт-Петербург, Ресей).

Аударма бағыттары: ағылшынша↔испанша, ағылшынша↔немісше, ағылшынша↔орысша, ағылшынша↔французша, испанша↔ағылшынша, испанша↔орысша, испанша↔французша, итальянша↔орысша, немісше↔ағылшынша, немісше↔орысша, немісше↔французша, орысша↔ағылшынша, орысша↔немісше, орысша↔французша, французша↔ағылшынша, французша↔испанша, французша↔немісше, французша↔орысша.

Интерфейс тілдері: ағылшын, орыс, француз, неміс, испан, итальян.

*PROMT Translation Office 2000/ XT* – еуропа тілдерінен орыс тіліне және керісінше орыс тілінен еуропа тілдеріне аударатын электрондық сөздік. *PROMT* сөздігінің көмегімен тек аударма ғана жасап қоймай, онда аударманы өңдеуге және бір уақытта барлық тілдердегі сөздіктермен жұмыс жүргізуге болады.

*PROMT-ma* жоғары сапалы аударма алу үшін нақты бір тақырыптағы мәтіндерді аударуға арналған, арнайы мамандыққа, пәнге негізделген әртүрлі сөздіктер құрастырылған, сонымен қатар қолданушының өзінің жеке сөздігін жасауға (құруға) мүмкіндігі бар.

*PROMT* жүйесінде әртүрлі сөздіктер бар, яғни әртүрлі салаларға (жеңіл өнеркәсіпке, ауыр өнеркәсіпке, сауда саласына, әртүрлі ғылым салаларына, техника саласына және тағы басқа) арналған сөздіктер.

Жүйе кез келген операциялық жүйелерде жұмыс істей алады. <http://www.translate.ru/> сайтында онлайндық тәртіпте аударма жасауға болады.

Бұл жүйе ұсынған қосымша мүмкіндіктерді атап өтсек, олар:

– қолданылатын сөздіктердегі сөздік қорды (қосымша, шылау, есімдік, омоним сөздер сияқты арнайы сөздерден басқасын) қайта өңдеу (редакциялау);

– сөздіктердің деңгейін анықтау (сөздіктен аударылатын сөздің бірнеше нұсқасы кездескен жағдайда);

– мәтіннің тақырыбын (тірек сөздер арқылы) автоматты түрде анықтау және арнайы сөздіктерді қолдану;

– бірнеше құрылымдарға аударма алгоритмдерін орнату (мысалы, ағылшынның “you” сөзін орыс “ты”, қазақ “сен” немесе орыс “вы”, қазақ “сіз” деп аудару сияқты);

– арнайы бірліктерді: файлдардың атауларын, электрондық мекенжайларды, сайт атауларын сол калпында аударуға (қайта беруге) мүмкіндігі бар;

– бағдарлама аудармайтын термин сөздерді, ерекше атауларды анықтайды;

– Translation Memory сақтау жүйесіндегі ассоциативті аударма қорларын пайдалану.

### ***L-MASTER – 98 электрондық сөздігі.***

Құрастырған: Trident Software, Ltd (Киев, Украина).

*L-MASTER – 98* бағдарламасы резидентті бағдарлама ретінде жүзеге асырылған, операциялар жүйесін іске қосқаннан кейін шығады, сонымен қатар Microsoft Word 97/2000 бағдарламаларында жеке құрал ретінде пайда болады.

*L-MASTER – 98* сөздігінің өзінің жеке интерфейстік терезесі жоқ, яғни бұл жүйені іске қосқанда индикатор белгі шығады, сол белгі арқылы жұмыс істейміз. *L-MASTER – 98* жүйесінің белгісіне барып тышқанның оң жағын бассaq, әртүрлі қызмет атқаратын (аударма жасау, орфографияны тексеру, қолданушылар сөздігі, анықтама және тағы басқа) контекстік мәзір шығады. Word бағдарламасында *L-MASTER – 98* электрондық сөздігі арқылы аударма жасасaq, ол аударма жеке Word терезесінде шығады және басқа да бағдарламаларда, мысалы, Note Pad, Internet Explorer бағдарламаларында да мәтін бөліктеріне аударма жасауға болады, мұндай жағдайларда аударма жұмыс терезесінде шығады. Бұл жағдайдың кемшілігі мәтінді көшіріп ала алмау.

Интерфейс тілдері (машинаға аударма жасау үшін енгізген тілдер): украин, ағылшын, орыс тілдері. Аударма бағыттары: сөздіктегі барлық тілдік жұптар.

### ***Magic Translator 3.0 электрондық сөздігі.***

Авторлық құқығы: 2000-2002 Zhang Duo. “Клиент-сервер” үлгісіндегі аудармашы.

Бұл клиенттік бөлімінде таңдалған аударма бағытына жауап беретін, негізгі қызметі мәтінді редакциялау, құрастыру, файлдармен жұмыс істеу (ашу, сақтау және т.б.) болып табылатын, жеке интерфейстік терезесі бар электрондық сөздік.

Интерфейс тілдері: ағылшын, француз, испан, неміс, итальян, португал және жапон.

Аударма барысында терезені жаппай-ақ, интерфейстік тілдер арасында бір тілден екінші тілге ауысуға болады.

Құрастырушылардың айтуы бойынша аударма жылдамдығы, егер 56 килобайттық модем құрылғысына қосылса, 1 минутта 3000 сөзді аударды.

***Translate Now! 1.6 (beta) электрондық сөздігі.***

Авторлық құқығы: 1991-2001 Magic Bit Software.

Өте қолайлы ағылшын-орыс және орыс-ағылшын тілдеріндегі контекстік аудармашы сөздік. Арнайы батырмалардың (клавиши) көмегімен аударма жасауға және сөзді әртүрлі мәтін терезелерінде (мысалы, Word, браузер веб-беттердегі терезелерде) қоюға мүмкіндік береді.

*Translate Now* жүйесімен жай электрондық сөздік ретінде жұмыс істеуге болады. Бұл сөздіктің нақты тіркелген нұсқасында сөздік қорды редакциялауға мүмкіндік бар және бұл электрондық оқыту жүйелері ретінде пайдаланылады. Бағдарламаның сөздік қорында әртүрлі салаға байланысты сөздіктер бар. Қажет болған жағдайда оларды көшіріп, орнатып алуға болады.

***Instant Messenger Translator Ver.1.0 электрондық сөздігі.***

Авторлық құқығы: Smark Link Corporation.

*Instant Messenger (IM) Translator* – онлайндық аудармашы. Бұл жүйе CQ және AIM сияқты бағдарламаларда бір тілден екінші тілге хабарландыру мәтіндерін және шағын көлемді мәтіндерді автоматты түрде аударуға арналған.

Интерфейс тілдері: ағылшын, орыс, француз, неміс, итальян, испан.

Қазіргі таңда аудармашы сегіз тілде жұмыс жүргізеді: ағылшынша↔французша, ағылшынша↔немісше, ағылшынша↔орысша, ағылшынша↔испанша, французша↔немісше, французша↔орысша, немісше↔орысша, итальянша↔орысша және керісінше аударма жасайды.

Аударма PROMT технологиясы негізінде жүзеге асады, аударма сапасы уақыт барысында тексерілген және атақты халықаралық мамандардың тексерісінен өтіп, дәлелденген.

***Translation Memory электрондық сөздігі.***

*Translation Memory* жүйесі, әдетте аудармашыға көмекші құрал ретінде қолданылады, МАНТ (аударманы бағдарлама құралдарын қолдану арқылы адамның өзі аудару) жүйесімен бірдей. Бұл сөздіктің мәні-маңызы жиі қолданылатын лексикалық және грамматикалық құрылымдары жинақталған аударма үлгілерін қайта қолдануында. Әдетте, мұндай ұсыныстарды мәтіндерді аудару барысында қолдана алмаймыз, тек жеке техникалық мәселелерді (бағдарламалау шеберлігінде, терминологиялық деректер қорын жасауда) шешуде немесе өндірістік құжат аудармаларында (өнім құрылысын, қолдану ережелерін, келісімшарттарды аударғанда) қолданылады.

*Translation Memory* (ТМ) – мамандардың орындаған аудармасын түпнұсқа-аударма мәтін ретінде сақтайтын деректер қоры.

*Translation Memory* аударма технологиясы аударылатын құжатты аударма қорында сақталған мәліметтермен салыстыруға негізделген.

*Translation Memory* технологиясы жинақтаушы (накопления) негізінде жұмыс істейді. *Translation Memory* жүйесінде аударма жасау барысында машинадан аударылып шығатын сөйлем мен оның аудармасы сақталады. *Translation Memory* сөздігіне аударылуға келіп түскен жаңа мәтінді аудару (өңдеу) барысында, машинадан аударылып шығатын сөйлемге пара-пар немесе ұқсас сөйлем табылса, онда ұқсас сөйлем аудармасы негізгі аудармамен бірге шығып, олардың ұқсастық (үлестері) пайыздары көрсетіледі. Сақталған мәтіннен ерекшеленетін жаңа сөздер мен фразалар жазылуына (мысалы, сақталған мәтін көк түспен жазылса, аударылып шығатын сөздер мен фразалар қызылмен жазылады деген сияқты) қарай анықталынады. Аудармашыға тек жаңа сөйлемдерді аударып, біртіндеп сәйкес келетіндерді қайта өңдеу қажет. *Translation Memory* сөздігінде әрбір өзгертулер мен жаңа аудармалар сақталады, ал нәтижесінде бір сөйлемді (әртүрлі жағдайларда кездесетін) екі рет аударудың қажеті жоқ.

«Бір рет жазып, бір рет сақтап, бір рет аударып – бірнеше рет қолданыңыз» деген мақсатта құрылған жүйе. Мұндай жұмыс әдісі қайталанатын және тақырыбы мен құрамы бойынша ұқсас мәтіндерді аудару барысында өте тиімді.

*Translation Memory* – бұл бүкіл орындалған аударма үдерісін есінде сақтайтын, үнемі өсіп отыратын (толығып отыратын) деректер қоры.

*Translation Memory* технологиясы (аз көлемді шығын барысында) жылдам және сапалы аудармамен қамтамасыз етеді. *Translation Memory* технологиясын қолдану, қайталанылатын сөйлемдері көп құжаттарды аудару үшін өте тиімді жүйе (экспорт-импорт келісімшарттарын аударуда, техникалық құжаттарды аударуда, өнімді түсіндіру құжаттарын аударуда және тағы басқа осы сияқты құжаттарды аударуда өте тиімді).

*Translation Memory* технологиясын қолдану кәсіби аудармашылардың еңбек өнімділігін арттырады, бұл аударылған сөйлемді қайта аудармауға көмектеседі.

*Translation Memory* электрондық сөздігін аударма жадысы (ағыл. translation memory *аударма жинақтаушысы*) – бұрын аударылған мәтін құралдарынан тұратын деректер қоры деп те атайды.

*Translation Memory* концепциясын, яғни аударманы жинақтаушы жүйелерді көпшілікке танымал бірнеше сөздіктер қолданады. Олар: Déjà Vu (Arti компаниясы), Omega T, Trados, Metatexis, Star Transit, Wordfast электрондық сөздіктер.

Машиналық аударманың үшінші түрі алдыңғыларға қарағанда біраз күрделі. Мұндай толық автоматтандырылған машиналық аударма жүйелерінің тиімділігі тіл мен ойлаудың қолданылу заңдылығының объективтілігіне байланысты болады.

Қазіргі таңда жиі қолданылатын толық автоматтандырылған машиналық аударма жүйелеріне мынадай электрондық сөздіктерді жатқызуға болады: Trados (Традос), Omega T, Déjà Vu (Дежавю) және тағы басқа электрондық сөздіктер. Машиналық аударма өте ірі және сұранысы көп жүйе. Толық автоматтандырылған аударма FАMТ жүйесіне арналған дербес ұсыныстар бар. Берілген топта өнімге қойылатын басты талап: аударма сапасы, үдеріс жылдамдығы және қолайлы интерфейс, сөздік қорды толықтыруды дамыту.

**Trados (Традос)** – автоматтандырылған аударма жүйесі алғашында 1992 жылы Trados GmbH неміс компаниясында құрастырылған. Trados электрондық сөздігі әлемдегі Translation Memory (ТМ аудармалардың жинақтаушы қоры –накопитель переводов) жүйесінің концепциясын қолданады. Trados GmbH компаниясы



1984 ж. Штутгарт қаласында ашылған. *Trados* электрондық сөздігінде әртүрлі форматтағы (мысалы, Microsoft Word құжаттарын, PowerPoint презентацияларын, HTML форматындағы мәтіндерді және Frame Maker, Inter Leaf құжаттарын және тағы басқа) мәтіндерді аударуға арналған бірнеше модульдер бар. *Trados* электрондық сөздігі бірнеше нұсқадан тұрады.

*Trados* жүйесіндегі жұмыстың мақсаты: Translation Memory концепциясы бойынша бұрын аударылған аудармаларды деректер қорынан сұрау арқылы қайта шығаруға болады.

*Trados* жүйесінің негізгі модульдері:

– *Trados Workbench* – Microsoft Word терезесінде аударма жасайды;

– *Tag Editor* – HTML, XML және тағы басқа форматтардағы мәтіндерді аударатын модуль;

– *Win Align* – бұрынғы аударылған екі тілдегі мәтіндер негізінде (аударма жадын) аударма орнын жасайтын модуль;

– *S-Tagger* – Frame Maker және Inter Leaf форматтарындағы құжаттарды аударуға арналған модуль;

– *T-Window* – PowerPoint форматындағы құжаттарды аударатын модуль;

– *ExtraTerm*, *Multi Term* – мәтіндегі түсініксіз сөздерге анықтама беретін түсіндірме сөздік, сөздік жасауға арналған модуль [19].

### ***Omega T электрондық сөздігі.***

*Omega T* – Java тілінде жазылған аударма жады бар, электрондық аударма жүйесі.

*Omega T* сөздігінің мүмкіндіктері: аударма барысында бұрын аударылған сөздерді, сөйлемдерді қолданады, әртүрлі сөздіктерді пайдаланады. Сөздік қордан мәнмәтінді тақырып бойынша іздейді және тірек сөздермен жұмыс істейді. GNU/Linux, MAC OS X және Microsoft Windows, Windows NT операциялық жүйелерінде *Omega T* сөздігімен жұмыс істеу үшін Java 1.4 нұсқасы (версиясы) қажет. *Omega T* сөздігі Unicode мәтіндік файлында, HTML/XHTML, Star office.org және Open Document (ODF), Doc Book файлдарында жұмыс істей алады. Ал *Omega T* сөздігі Microsoft Office (Word, Excel және PowerPoint) мәтіндік редакторында тікелей жұмыс істей алмайды. Microsoft Office мәтіндік редакторында жұмыс істеу үшін файлдарды Open Document файлына көшіре алатын Open Office.org бағдарламасын қолдану қажет.

*Omega T* жүйесінің ең кең таралған нұсқасы *Omega T 1.6* сөздігі (19.11.2006), ал ең соңғы шыққан нұсқасы *Omega T 1.7* сөздігі (28.11.2007) [19].

*Déjà Vu* (Дежавю) жүйесі – испандық Artі компаниясының өнімі 1990 жылдың екінші жартысында жасалған. *Déjà Vu Translation Memory* – аударма жадысымен жабдықталған, көпшілікке белгілі, танымал электрондық сөздік.

Машиналық аударма жүйесінің қызмет етуіне машиналық аударма жүйелерінің лингвистикалық, бағдарламалық және ақпараттық жабдықталуы тиіс. Лингвистикалық жағынан жабдықталуына сөздіктер, сөз тіркестері, қосымшалардың кестесі, грамматикалық ережелер қоры және т.б. жатады. Бағдарламалық қамтамасыз етілуіне аударма бағдарламалары, ақпараттыққа пәндік салаға қатысты экстралингвистикалық білім қорларын жатқызамыз.

### Семинар тапсырмалары

1. Машиналық аударма дегеніміз не? Оның қандай түрлері бар? Машиналық аударма, автоматты аударма, автоматтандырылған аударма термин тіркестерін түсіндіріп беріңіз. Қайсы дұрыс?
2. Машиналық аударма кезеңдерін сипаттап беріңіз.
3. Машиналық аударма адам қандай рөлге ие болады? Адамның қатысы қажет пе? Дәлелденіз.
4. Бірнеше мәтінді машинаның көмегімен бір тілден екінші тілге аударып, өңденіз. Қандай қиындықтар кездеседі?
5. Автоматтандырылған машиналық аударма жүйелерін сипаттап беріңіз. Оларды неліктен электрондық сөздік деп атайды?
6. Елімізде қолданысқа ие машиналық аударма жүйелерін атаңыз.

### Әдебиеттер:

*Негізгі:*

1. Марчук Ю.Н. Автоматический перевод // Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998. – С. 15.
2. Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практ. руководство для студ., аспирантов, препод.-филологов. – М.: Флинта: Наука, 2007.
3. Марчук Ю.Н. Модели перевода. – М.: Академия, 2010.
4. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
5. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. СПб.: Книжный Дом, 2007.

6. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
7. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.
8. Марчук Ю.Н. Автоматический перевод // Большой энциклопедический словарь. Языкознание. М.: Большая Российская энциклопедия, 1998. – С. 15.
9. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие для студ., аспирантов, преподавателей-филологов. 2-е изд., испр. – М.: Флинта: Наука, 2007.
10. Башмаков И.А., Башмаков А.И. Интеллектуальные информационные системы. – М.: МГТУ им. Н.Э. Баумана, 2005.
11. Бонч-Осмоловская А.А., Толдова С.Ю. Разрешение анафоры // Фонд знаний «Ломоносов». – М., 2011. [www.lomonosov-fund.ru/enc/ru/encyclopedia:0127469:article](http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127469:article) (қаралған күн: 28.02.2012).

*Қосымша:*

1. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации: учебно-метод. пособие. – М., 2009.
2. Русский ассоциативный словарь / Ю.Н. Караулов, Ю.А. Сорокин, Е.Ф. Тарасов. Кн. 5: Прямой словарь: от стимула к реакции. – М.: ИРЯ РАН, 1998.
3. Селегей В. Электронные словари и компьютерная лексикография // Ассоциация лексикографов Lingvo. [www.lingvoda.ru/transforum/articles/selegey\\_a1.asp](http://www.lingvoda.ru/transforum/articles/selegey_a1.asp) (қаралған күн: 28.02.2012).

## **10-11 дәріс. Компьютерлік лексикография**

**Қарастырылатын мәселелер:** *компьютерлік лексикография түсініктемесі; электрондық сөздік; сөздік мақала құрамы; электрондық сөздік түрлері; электрондық сөздіктің ұтымдылығы; компьютерлік лексикографияның бағыты.*

**Тірек сөз бен сөз тіркестері:** *компьютерлік лексикография, электрондық сөздік, сөздік мақала, электрондық сөздік, автоматты сөздік, машиналық аударма, автоматты аударма, автоматты рефераттау, дәстүрлі сөздік, өңдеу бағдарламалар, операциялық жүйе, аудармашы.*

**Компьютерлік лексикография** дегеніміз компьютерлік сөздік шығаруға бағытталған, лингвистиканың қосарлама бір бөлігін, деректердің лингвистикалық ақпараттар базасын және лексикографиялық жұмыстың зерттелу бағытын қарастыратын пән.

Дәстүрлі және компьютерлік лексикографияның негізгі мақсаты – сөздіктің құрылымын және сөздік мақаланың аймағын анықтау, сонымен қоса әртүрлі сөздік түрлерін белгілі бір қағидаларға сай жасап шығару.

**Электрондық сөздік** – қажетті сөздерді тез аударуға, сөздіктен іздеуге және аударма бағытын, мысалы, ағылшын-орыс тілдерінен, орыс-ағылшын тілдеріне өзгертуге мүмкіндік беретін арнайы сөздік мақалаларынан, сөздіктерден құралған компьютерлік деректер қоры.

Электрондық сөздіктерді автоматты, машиналық немесе компьютерлік деп те атайды.

Электрондық сөздіктер оларға негізделген машиналық аудармаларға қарағанда едәуір ерте пайда болған. Сондықтан қазіргі ақпараттық технологиялар дамыған заманда өте қарапайым электрондық сөздіктерден бастап, бірнеше лексикалық деректер қорын бір бағдарлама жүйесіне біріктірген, арнайы тақырыптарға арналған электрондық сөздіктерге дейін кездестіруге болады. Электрондық сөздіктерді қолданушылар өздерінің мүмкіндіктеріне, қолдану мақсатына байланысты таңдап алады. Электрондық сөздіктер әртүрлі мақсатта қолданылады. Оларды шет тілін оқу барысында оқу құралы ретінде, алуан түрлі ақпараттарды, мәтіндерді аудару кезінде анықтағыш, аударма сөздік ретінде көмек бере алады [1, 55-б.].

Өзіндік ерекшелігіне байланысты автоматтық сөздіктің екі түрі айқындалады. Олар: *арнайы қолданушы адамға арналған автоматты сөздік және мәтінді түзететін аудару бағдарламаларға арналған автоматты сөздік.*

Арнайы қолданушы адамға арналған автоматтық сөздік көбіне белгілі қарапайым сөздіктердің компьютерлік нұсқасы болып табылады [2, 86-б.]. Мысалы:

- ағылшын тілінің оксфордтық сөздігі;
- ағылшын тілінің автоматтық анықтама сөздігі;
- «Жаңа үлкен ағылшын-орыс тілі сөздігі» сөздігінің автоматты нұсқасы;
- Ожеговтың онлайн сөздігі, т.б.

Машиналық аударма жүйесіне арналған, автоматты рефераттауға арналған, мағлұматты іздеуге арналған және т.б. автоматтық сөздіктер сөздік мақаласының интерфейсінен және құрылымы-

нан әлдеқайда ерекшеленеді. Олардың құрылымдық ерекшеліктері, сөздік құжаттарының бағытын өлшеу өздерінің арнайы сәйкестендірілген бағдарламалармен белгіленеді. *Мұндай сөздік* бірден жүзге дейін сөздік мақалаларын мазмұндай алады. Лексикографиялық сипаттаманың ауданы шектен тыс сан түрлі: морфологиялық, лексикалық, синтаксистік, семантикалық және т.б.

Дәстүрлі сөздік, негізінен, келесі компоненттерден тұрады [1, 86-б.]:

– сөздікті қолдану қағидаларын түсіндіретін және толық сөздік мақала құрылымы туралы мағлұмат беретін кіріспе;

– сөз бірліктерінен тұратын сөздік: морфемдер, лексемалар, сөз үлгілері және сөз тіркестерінен; мұндай сөздік бірліктерінің әрқайсысы өзіне сәйкес сипаттамасымен сөздік мақаланы ұсынады;

– нұсқаулар (индекстер);

– ақпараттар тізімі;

– шартты қысқарту мен әліпби тізімі [3, 75 – 76-б.].

Электрондық сөздік пен қарапайым кітап бетіндегі сөздіктің айырмашылығы – мультимедиялығы және гипермәтінділігі. Бұл баспадан шыққан сөздікке қарағанда электрондық сөздікке көбірек тән. Осылайша, гиперсілтемелер сөздік мақаласының әрбір элементінде немесе сөздіктің *мәзір* бағдарламасының тармағына жүктелген болуы мүмкін. Осыған орай қолданушыға іздеу және өзіне керек сөздік мәліметтеріне жылдам ауысуға, берілген сөзге антоним мен синоним табуға қосымша мүмкіндік береді.

Компьютерлік сөздіктер автоматты түрде деректерді өңдеу тәсілін қолданады және сөздік бірліктерін іздестіру арқылы мәтіннің бір бөлігі негізінде жасалады. Ол үшін сөздік мәтінін автоматты түрде қалыпқа келтіретін, сөздік ақпараттарын сақтайтын және оны өңдейтін дерек қоры, компьютерлік картотекалар, мәтінді өңдеу бағдарламалары сияқты арнайы бағдарламалар керек. Электрондық сөздікті жасап шығару келесі кезеңдерге сәйкес жүреді:

– мәтінді құрастыру және сөзтізбенің қатарлас құрылымы;

– мысал тұрқының автоматты түрде қалыптасуы;

– сөздік мақаласын жазу;

– мақала сөздігін дерекқорға енгізу;

– сөздік мақалаларын дерекқорда редакциялау;

– дерекқордағы мәтін қатесін түзету;

– сөздік мәтіннің туындауы мен түпнұсқалы пішінде қалыптастыру;

– сөздік басылымы.

Әрине, электронды сөздік жасау үдерісінің көрсетілген сипаттамасының түріне, зерттеу принциптеріне және басқа да факторларға байланысты түзету енгізуге болады. Бірақ барлық жағдайда да компьютерді және компьютерлік лексикографиядағы дайын мәтінді қолдану электрондық сөздік жасау үдерісінде уақытты үнемдеуге көмектеседі.

Электрондық сөздіктерді негізгі техникалық және пайдалану сипатына қарай операциялық жүйелерді қолдану, жұмыс істеу көлемі, сөздіктердің қосылу саны, сөздік қорын кеңейту мүмкіндіктері, аударма тәртiбi сияқты қағидалар бойынша жіктеуге болады [4, 59 – 74-б.].

Қолданылатын операциялық жүйелер ең қарапайым электрондық сөздіктер [Dic] 2.21 нұсқасынан бастап 3.30 нұсқасына дейінгі MS-DOS операциялық жүйелерінде жұмыс істейді, яғни қарапайым электрондық сөздіктерді кез келген компьютерлерде қолдана аламыз. Қызметі жағынан көп функциялы, көп терезелі күрделі электрондық сөздіктер Windows 3.11, Windows NT, Windows 95, 98, 2000 және тағы басқа операциялық жүйелерде жұмыс істейді. Электрондық сөздіктердің жылдам жұмыс істеуіне оперативті жады (оперативная память) көлемінің 8 мегабайттан кем болмауы қажет, бірақ қазіргі уақыттағы шығып жатқан компьютерлердің оперативті жадылары бұрынғыға қарағанда күннен-күнге бірнеше есеге арту үстінде.

Жүктеу тәсілі бойынша (по способу загрузки) электрондық сөздіктерді резидентті емес және резидентті деп бөлуге болады. Алғашқы қатарға тек бір ортада жұмыс істейтін, басқа терезелерден, мысалы, мәтіндік редакторларды шақыра алмайтын қарапайым электрондық сөздіктер (DIC сөзбе-сөз сөздігі) жатады. Көп жағдайда қарапайым электрондық сөздіктер автоматты аударма ретінде қызмет атқарады. Екінші қатардағы электрондық сөздіктер өздерінің жұмыстарын компьютердің оперативті жадына жүктейді (мысалы, DOS оперативті жүйесіне арналған “Lingvo”). Бұл сөздіктер компьютердің жұмыс істеуіне байланысты, кез келген уақытта арнайы батырмалардың, операцияны шақыру, ашу бұйрықтарын басу арқылы мәтіндік редакторларды бірден

аша аламыз. Мұндай электрондық сөздіктер интерактивті аудармашылар қызметін атқарады.

Электрондық сөздіктерге қосылған сөздік қорлар, сөздіктер санына байланысты да бөлуге болады. Электрондық сөздіктердің алғашқы нұсқалары тек бір ғана сөздікпен жұмыс істеген, ал қазіргілердің жүзеге асыра алатын мүмкіндіктері көп, мысалы, Lingvo – электрондық сөздіктер жүйесі қандай оперативті жүйеде орнатылғанына қарамастан, бірнеше ондаған сөздік қорларды қосып, кейінгі шыққан сөздіктерді орната алады.

Сөздіктер қорын кеңейту мүмкіндіктері бойынша электрондық сөздіктердің алғашқы нұсқаларында сөздік қорын кеңейтетін мүмкіндіктер болмады, ал олардың қазіргі нұсқаларында, мысалы, Lingvo электрондық сөздігінің Lingvo 4.6 нұсқасынан бастап, жаңа нұсқаларында қолданушы өзінің жеке сөздерін сақтау үшін және сөздіктер қорын кеңейту үшін арнайы сөздіктер түрлерін құрастыра алады (яғни электрондық сөздіктердің сөздік қорын кеңейтетін, жаңадан құрастыратын мүмкіндіктер бар).

Аудару тәртібі бойынша (по режиму перевода) аударманың екі негізгі тәртібін ажыратуға болады: автоматты (сөзбе-сөз, жолма-жол) және интерактивті («сұрау-жауап беру»).

Алғашқы жағдайдағы электрондық сөздік аударылуға қажетті мәтіндік құжатты көрсетілген атауы бойынша немесе аударылуға арналған толық мәтін бойынша шақырады. Содан кейін аудармашының қатысынсыз әрбір жеке сөздің аудармасын монитор экранында көрсетеді немесе әрбір аударылған сөздің орыс тіліндегі баламасын көрсетіп, электрондық сөздікте жоқ сөздердің орнына “?” (сұрақ) белгісін береді және аударылатын сөздер мен аударылатын мәтіндерге төте аударма жасалынады. Мұндай тәртіп сөзбе-сөз немесе жолма-жол аударма тәсілі (подстрочный) деп аталады. Осындай электрондық сөздіктердің сөздік қоры өте шектелген, аударма сапасы орташа.

Екінші тәртіптегі электрондық сөздіктер компьютердің оперативті жадына сөздіктің резидентті бөлігінің жұмысын жүктегеннен кейін, қолданушы, мысалы, мәтіндік редактордан не сөздік сыртында жұмыс істеп отырып, түпнұсқа тіліндегі белгісіз бір сөзді пернетақтамен (клавиатура) енгізіп, одан кейін тышқанның көмегімен ерекшелеп алып, “арнайы батырмаларды” басып электрондық сөздіктердің жұмысын жылдамдатып, түпнұсқа мә-

тінді, мысалы, орыс тілді аудармаға ауыстыра алады. Орта есеппен аударма мәтінді шығару уақыты – 0,2 секунд, бұл аудармашы жұмысын жылдамдатады. Өзінің мүмкіндіктерін кеңейтетін және бір мезгілде бірнеше сөздік қорды қоса алатын нақты осындай электрондық сөздіктер әртүрлі кәсіби деңгейдегі аудармашылар жұмысын жеңілдетіп, жылдамдата алады.

Компьютерде машиналық аударманы біз әртүрлі электрондық аудармашыларды, электрондық сөздіктерді қолдана отырып, жүзеге асырамыз. Ешқандай компьютер машиналық аударманы электрондық сөздіксіз жасай алмайды. Компьютер арқылы машиналық аударманы жасау үшін, ең алдымен, арнайы аударманы жүзеге асыруға арналған бағдарлама болуы қажет, яғни компьютерге аударманы жасайтын бағдарламаларды, электрондық сөздіктерді орнату керек.

#### ***ABBY Lingvo электрондық сөздігі.***

*ABBY Lingvo* электрондық сөздіктер жүйесі ресейлік *ABBY* компаниясында жасалған. Сөздіктің бірнеше нұсқасы бар. Ең соңғы 12-нұсқасы 2006 жылы 12 желтоқсанда шыққан. Қазіргі таңда *ABBY lingvo* сөздігінің 12-нұсқасы көп сөздіктермен қамтамасыз етілген, толық жетілген электрондық сөздіктердің бірі деп айтуға болады. Бағдарламаның көптілді нұсқасында әртүрлі тақырып бойынша (оның ішінде сыра мен шарап жасау шаруашылығына арналған) жиыны 128 сөздік бар.

*ABBY Lingvo* сөздігі толық мәтіндік аударма жасай алмайды, бірақ үлкен көлемді мәтіндерді сөзбе-сөз аударғанда өте пайдалы. Бұл – сөздіктің қазіргі ең қызықты қызметтерінің бірі, ол аударылатын сөзге тышқанды апарып басқанда аудармамен қалқып шығатын терезе. Сөздіктің ағылшын, француз және неміс тілдеріндегі бөліктерінде кәсіби дикторлар (тілді жетік меңгерген) сөздерді дыбыстап (айтады) оқиды (үндендірген), бұл тұрғыдан алғанда сөздік тілді жаңадан үйренушілерге өте қолайлы. Осы сөздік құрамына жаңа сөзді есте сақтауға көмектесетін, сөздерді оқитын *Lingvo Tutor* модулінде кірген.

128 кәсіби сөздіктен басқа *ABBY* ұйымының қызметкерлерінің лексикографиялық жұмыстарының қорытындысын, қағаз және электрондық сөздіктердің кеңейтілген түрлерін тегін қолдану мүмкіндіктері бар. Компанияның және сөздіктің сайты: <http://www.abbyonline.com/ru>, <http://www.lingvo-online.ru/ru>



*АВВУ* *Lingvo* электрондық сөздігі: орыс, ағылшын, неміс, француз, испан, итальян, қытай және украин тілдеріндегі мәтіндерді аударады.

### ***Lingvo* электрондық сөздігі.**

*Lingvo* электрондық сөздігінің бірнеше нұсқасы бар.

Бұл сөздіктің *Lingvo 7.0* нұсқасы қолданушыға өте ыңғайлы кәсіби сөздік. *Lingvo 7.0* сөздігінде кез келген сөзге грамматикалық түсініктеме беріледі, қиын сөздерді дауыстап оқиды (айтады) және емлені, яғни дұрыс жазуды тексереді, сонымен қатар қолданушы өзінің жеке сөздігін жасай алады. 18 жалпы және арнайы мамандыққа арналған *Lingvo 7.0* сөздігінде 1,2 миллионнан астам сөз бен сөз тіркесі бар.

*Lingvo* сөздігін іске қосқанда экранда бас терезе пайда болады. Енгізу жолында қолданушы қажетті сөзді теріп, жазуы керек, ал одан кейін бағдарлама оған сәйкес сөзді іздейтін болады. Енгізу бағдарламасын немесе «мәтінді аудару» пиктограммасын басу арқылы қолданушы таңдап алынған (іздеу барысында табылған) сөздің сөздіктегі аудармалары берілген (карточка) терезені (окно карточки) көре аламыз. *Lingvo* сөздігінде мәтіндегі бір немесе бірнеше сөзді қатар белгілеп алып, Ctrl+Ins+Ins батырмасын басу арқылы не болмаса сөздерді енгізу жолына (drag-and-drap) қойып, аудару белгісін басу арқылы бірден аударуға болады. Өңделетін мәтінге аударманы қою үшін карточка терезедегі аударманы белгілеп, Ctrl+Ins батырмаларын басып, мәтін редакторына (Ms. Word, Ms. WordPad және тағы да басқа редакторларға) барып «енгізу» (вставить) бұйрығын басу қажет. Қажетті аударма мәтін редакторында шығады.

Ағылшын тілінен орыс тіліне аудару барысында әрқашанда жеке аударылатын сөздерді және сөз тіркесі құрамында тұрып бір тұтас мағынаны білдіретін сөздерді де кездестіреміз. Бұл мәселені шешуге арнайы сөздіктері, сөздік мақалалары бар, аударылатын мәтіндегі тұрақты сөз тіркестерін табуға мүмкіндік беретін “мәтінді жол бойынша аудару” (“перевести текст из строки”) қызметі көмектеседі. Аударылмай қалған сөйлемдерді, сөйлем бөліктерін (фрагменттерді), толық мәтіндік іздеу қызметінің көмегімен табуға болады. Орыс тілінен ағылшын тіліне аудару барысында сөз тіркестері мен грамматикалық құрылымды қиындықсыз белгілеп аламыз, ал берілген сөз тіркесі сөздікте жоқ болған жағдайда бірден толық мәтіндік іздеу қызметін іске қосуға

болады. Бұл іздеудің қорытындысынан қажетті сөйленістің, сөйлемнің нақты мысалдарда қалай аударылатынын көруге болады.

*Lingvo* электрондық сөздігінің басты ерекшеліктері:

- сөздік қоры 1,2 миллион сөз бен сөз тіркестерінен тұрады;
- 18 жалпы және арнайы мамандыққа арналған сөздіктер бар;
- қазіргі лексика, яғни жаңа сөздер толық қамтылған;
- кез келген операциялық жүйеде жұмыс істей алады (Linux/Unix және Windows операциялық жүйелерінің барлық нұсқаларында);
- іздеу жүйесі дамыған;
- ағылшын тіліндегі 5 мың сөздерді дауыстап оқиды;
- қолданушының өзіне жеке сөздігін жасауға мүмкіндігі бар;
- сөздердің түсіндірмесін беріп, оның қолданылу жолдарын түсіндіреді;
- арнайы мамандыққа арналған сөздіктердің жаңа толыққан нұсқасы бар;
- сөздіктің кейбір бөлігін интернет желісінде отырып тегін қолдануға болады.

***MultiLex электрондық сөздігі.***

*MultiLex 3.5* көпшілікке белгілі, құрамында әртүрлі сөздіктерді қамтитын бағдарлама. “*MultiLex 3.5. Ағылшынша*” электрондық сөздігінде ағылшынша-орысша және ағылшынша-орысша сөздіктерінің әртүрлі нұсқалары бар (В.К. Мюллердің «*Жаңа ағылшынша сөздігі*, О.С. Ахманованың *Ағылшынша-орысша және Орысша-ағылшынша сөздігі*, А.И. Смирницкийдің редакциялауымен шыққан *Орысша-ағылшынша сөздік*).

*MultiLex 3.5* сөздігінің құрамына техника, медицина, биология, экономика, заң салаларына арналған сөздіктерді жасап, құрастырып қосу жоспарлануда.

*MultiLex 3.5* электрондық сөздігінің ерекшелігі:

- қарапайым, қолдануға ыңғайлы;
- сөздерді дауыстап оқиды;
- жылдам теру қызметі бар, сөзді терген кезде сізге қажетті сөзге ұқсас сөздер тізімі шығады, одан сіз өзіңізге қажетті сөздерді толық термей-ақ, бірден таңдай аласыз;
- сөз немесе сөйлемнің аудармасының қорытындысын Windows-тан алмасу (ауысым) буфері немесе drag-and-drop бұйрығы арқылы ауыстырып алуға болады;
- қолданушылар өз сөздіктерін жасай алады.

*MultiLex* электрондық интернет желісінде онлайндық нұсқасы бар.

*MultiLex* сөздігінде кәсіби лингвистер мен аудармашылар құрастырған жалпы 40-тан астам тақырыптық және түсіндірме сөздіктер бар. Сөздің мағынасын, транскрипциясын және қолдану тәсілдерін нақты түсіндіретін *MultiLex* сөздіктерін қолдана отырып, кез келген күрделі мәтінді аударуға болады. *MultiLex* сөздігінің құрамына арнайы сөздерді аударуға арналған құрылыс, экономика, банк саласы, компьютер, құқық, техника, медицина және тағы басқалар сияқты салаларда қолданылатын тақырыптық сөздіктері кіреді. *MultiLex* сөздігін кез келген уақытта жеке компьютерге (ЖК) және ұялы телефонға көшіріп алуға болады.

Аударма бағыттары: ағылшынша↔орысша, орысша↔ағылшынша, французша↔орысша, орысша↔французша, немісше↔орысша, орысша↔немісше, итальянша↔орысша, орысша↔итальянша, португалша↔орысша, орысша↔португалша, испанша↔орысша, орысша↔испанша.

*MultiLex* сөздігінің электронды адресі: <http://www.ontrans.ru/articles/17/18/38>, <http://online.multilex.ru/>

#### ***Multitran* электрондық сөздігі.**

Автоматтандырылған *Multitran* сөздігі бұл орыс, ағылшын, неміс, француз, испан, итальян, нидерланд, латыш, эстон және жапон тілдерінен аударма жасайтын электрондық сөздіктер жүйесі.

*Multitran* электрондық сөздігінің бағдарламасын құрастырушы – А. Поминов (Ресей, Мәскеу). Бұл – қазіргі уақыттағы ең толық, көпшілікке танымал автоматтандырылған онлайндық сөздіктердің бірі. Сөздіктің ағылшын тілінен орыс, неміс тілінен орыс және француз тілінен орыс тіліне аударатын бөліктері жетік толықтырылған, ал жапон тілінен орыс тіліне аударатын бөлігі жеткілікті түрде толықпаған. Күн сайын Multitran.ru сайтына отыз мыңға жуық адам кіреді, ал сервер 1 миллионнан астам іздеу сұраныстарын өңдеп, оларға жауап береді. *Multitran* электрондық сөздігінде Мультитран форумдары үнемі жұмыс істейді, онда әрбір сөздікті қолданушы өз сауалдарын қойып, аудармашылар ұжымымен хабарласа алады. Әрбір кезеңде Multitran қолданушыларымен офф-лайн жүйесінде форумдар өткізіліп тұрады.

*Multitran* сөздігінің сипаты:

– сөздікте барлық тілдерде 5 миллионнан астам термин бар;

- сөздіктің әрбір тілдік жұбының терминдері аударманың екі бағытында да қолданылады;
- сөздікте 800-ден аса пәндік сала бар;
- сөздікті қолданушылар өз бетімен толықтыра алады;
- сөздіктің онлайндық нұсқасын 1000-нан астам аудармашылар тікелей сайтта толықтырады;
- сөз тіркестерін алфавит, грамматика, фраза бойынша іздейтін жүйелер бар;
- мәтінді жол (подстрочное) бойынша аударды;
- сұрау мәтіндегі тұрақты сөз тіркестерін автоматты түрде іздейді;
- бір мезгілде сөздіктің бірнеше бөліктерінен, мысалы, ағылшын, орыс тілдеріндегі сөйлемдер қорынан іздеу жүргізе алады.

*Multitran* сөздігін компьютерге орнату арқылы, интернет желісінде отырып, War сайты арқылы ұялы телефонға, Gnt/Linux жүйесіндегі персоналды компьютерде қолдана аламыз.

*Multitran* сөздігінің нұсқалары Windows жүйесіне, интернет желісіне, War сайтына, Gnt/Linux жүйесіне және тағы басқа жүйелерге арналған.

*Multitran* сөздігінің артықшылығы мен кемшіліктері:

- сөздік қоры үлкен сөздіктерді тізбектеу арқылы құрастырылған, онда сөздердің алынған аудармасы деректер қорында сақталады;
- аударылатын сөздің бірнеше мағынасы беріледі. Мұндай жағдайда тілді жаңадан бастап үйреніп жатқан қолданушыларға аударма нұсқаларының ішінен ең негізгісін таңдап алу қиын және *multitran* сөздігі тілді үйрететін сөздік емес;
- әртүрлі білім саласына қатысты аударманы іздеудің қорытындысында бір аударманы үнемі бере береді;
- аударманың көп нұсқасын береді, оның ішінде қолдану аясы тар, яғни аз таралған сөздер де бар;
- сөздік қордың сапасын жақсарту мүмкіндіктері де қарастырылған. Сөздіктің әрбір тіркелген қолданушысы өзінің сөздіктен кездестірген қателерін сөздікті құрастырушыларға хабарлай алады.

*Multitran* сөздігінің электрондық адресі: <http://www.multitran.ru/c/m.exe>

***Polyglossum* электрондық сөздігі.**

*Polyglossum* сөздігін CD-ROM дискісіне “ЭТС” баспасы басып шығарған.

*Polyglossum* сөздігінің бірнеше нұсқасы бар. *Polyglossum II* сөздігінде 1,5 миллион сөз бар, ағылшын тілінен орыс тіліне және керісінше аударарды. Біріншіден, бір дискіде ағылшынша-орысша 8 сөздіктен тұрады. Олар: оқулық сөздіктер, экономика, банк, қаржы және несие терминдеріне арналған, шаруашылыққа, жалпы техникалық, математикалық, медициналық, бизнес (сауда) салаларына арналған сөздіктер. Екіншіден, көлемі бойынша ерекшеленеді, сөздіктің жалпы көлемі 1,5 миллион сөз (әрбір сөздікте 30 мыңнан 600 мыңға жуық сөз бар). Үшіншіден, сөздіктер жинағы екі түрде шығарылған. Олардың біреуі – Dos, Windows операциялық жүйелерімен жұмыс істейтін компьютерлерге арналған, екіншісі – Macintosh компьютерлеріне арналған. Төртіншіден, жинақтағы барлық сөздіктер лицензиялы және сөздіктердегі материалдардың барлығы Ресейдегі аудармашылар орталығы комиссиясының тексеруінен өткен.

*Polyglossum* сөздер мен сөз тіркестерін аударарды. Сөздіктің қоры күннен-күнге кеңейіп келеді. Қазіргі кезде *Polyglossum* электрондық сөздігіндегі тілдік жұптар да кеңейтілген. Қазір орыс, ағылшын тілдерінен басқа неміс, француз, испан, голланд, швед және фин тілдеріне аударма жасалынады.

#### ***Context электрондық сөздігі.***

*Context* электрондық сөздігінің бірнеше нұсқасы бар. *Context 4.0* – бағдарламасы дамыған, сөздіктер құрамы, әсіресе жалпы лексикалық және арнайы мамандыққа арналған сөздіктер қоры кеңейтілген электрондық сөздіктер жүйесі. *Context* сөздігінің бір ерекшелігі: аударғанда тілдердің морфологиялық сипатын да ескереді. Осының арқасында *Context* электрондық сөздігі кез келген грамматикалық формадағы сөз бен сөз тіркесін аударар алады. *Context* сөздігінде ағылшын тілінен орыс тіліне және керісінше аударатын, яғни орыс тілінен ағылшын тіліне аударатын сөздіктер толық берілген. *Context* сөздігінің жаңа нұсқасы ағылшын тілінен француз, неміс, испан, итальян, португал, серб, харват тілдеріне аударатын сөздіктермен толықтырылған. *Context* сөздіктері екі жақты болып есептелінеді. *Context* электрондық сөздігінің бағдарламасы бір тілден басқа тілге және ешқандай кедергісіз сол тілдерден керісінше аударма жасайды. Сөздің аудармасын жүйеге қосылған барлық сөздіктерден немесе нақты бір сөздіктен ғана іздеуге болады. *Context* сөздігінде жұмыс істегенде өзіне

қажетті сөзді немесе сөз тіркесін арнайы енгізу жолында (в специальном поле ввода) теруге, жазуға болады.

*Context* сөздігімен Windows операциялық жүйесінде жұмыс істеу өте ыңғайлы. Онда аударма drag-and-drap әдісі арқылы немесе буфер алмасу тәсілімен жүзеге асады. Ms.Word редакторында жұмыс істейтін қолданушыларға *Context* толық сөздігін мәтін редакторының (Ms.Word) өзінен ашуға мүмкіндік жасалған. Ол үшін Ms.Word құралдар үстеліндегі (панель инструментов) “контекст” батырмасын басу қажет. Мұнда мәтіндегі сөзді және сөз тіркестерін ерекшелеп алудың қажеті жоқ. *Context* сөздіктері қолданушылардың тапсырған сөздіктерімен толықтырылады. Егер қолданушы бірнеше сөздікпен қатар бағдарламаны сатып алса, ол кез келген уақытта өзіне қажетті басқа сөздіктерді сатып ала алады.

*Context* сөздігінің 4-нұсқасында бірнеше жаңа мүмкіндіктер бар. Сөздіктен жеке сөз тіркестерін іздей аламыз, сөздікке жылдам теру, жазу (Fast Typing) қызметтері қосылған. Қолданушы сөзді енгізу барысында таңбасы жағынан ұқсас және оның алдында енгізген сөздіктер ішінен көмек (подсказки) ала алады. Одан кейін қолданушыға енгізетін сөзін тізімнен таңдап алуына немесе өзі теруіне (жазуына) болады. Автоматты түрде табумен қатар әртүрлі тілдердегі сөздіктерін біріктіріп, жұмыс жүргізу қызметі мен тілді таңдау қызметі қосылған [22].

#### ***Wiktionary электрондық сөздігі.***

*Wiktionary* – түсіндірме, көптілді тезаурустық сөздіктердің бірігуінен құралған сөздік. *Wiktionary* жеке сөздерді аударарды, сөздің аудармасымен қатар ол сөздің түсіндірмесін, грамматикасын сипаттайды. Сонымен қатар мәтіндегі сөздердің семантикалық байланысы жайында да мәліметтер (ақпараттар) береді.

#### ***Start Dict электрондық сөздігі.***

*Start Dict* – GNOME жобасының бөлігі болып есептеледі. Linux/Unix және Windows операциялық жүйелерінде жұмыс істейді. *Start Dict* сөздігі сөздерді аударарды және дыбыстайды, яғни дауыстап оқиды [22].

2006 жылы қыркүйек айында жобаның негізгі барлық сөздіктері қосылған онлайндық нұсқасы интернетке қосылды.

Сөздіктің мүмкіндектері:

– шаблон бойынша іздеу, «\*» және «?» шаблон ретінде сөздерді енгізуге болады;

– индекснің көмегінсіз сөздіктен сөзді іздеуге арналған толық мәтіндік іздеу.

– ерекше қызметтерінің бірі: белгіленген сөздерді сканерлеу; бағытқа байланысты сөзді белгілеп алып, батырманы басқаннан кейін оның аудармасы басқа (қалқып шығатын *всплывающее окно*) терезеде көрінеді;

– сөздіктерді басқару, қажетті сөздіктерді қосу және сұрау ба-рысында олардың қолданылуын тәртіп бойынша жүзеге асыру;

– интернеттен әртүрлі онлайн сөздіктерін іздеу;

– сөздерді айту, дауыстап оқу.

Сөздіктің интернет желісіндегі электрондық адресі: <http://rutracker.org>.

### ***Sozдіk қазақша-орысша электрондық сөздік.***

*Sozдіk* – қазақ тілінен орыс тіліне және керісінше орыс тілінен қазақ тіліне аударатын электрондық сөздік, интернет желісі пайда болғаннан кейін жасалған. Қазақстан мемлекеті өзінің егемендігін алғаннан кейін және қазақ тілінің маңызы күнделікті тұрмыста, кәсіби жұмыста арта бастағаннан кейін, бұл жобаның өзектілігі одан әрі көтеріле бастады. Сөздік негізінде қазақ тіліне деген сұраныстың артуы нәтижесінде пайда болған. Бұл сөздік арқылы интернет желісінде отырып, өзімізге қажетті сөзді немесе сөз тіркесін аудараламыз. Аударма жылдамдығы өте жылдам. Сөздік қорында қазақ тілінде 98273 сөз, орыс тілінде 104610 сөз және 4161 аудиосөз енгізілген. Сөздік қорын кеңейту жұмыстары тұрақты түрде жүргізіліп жатыр және сөздікке экономика, техника, мұнай, газ салаларының терминдерін енгізу ең маңызды іс болып табылады. Сөздік Online жүйесіне жатады. Бұл сөздік арқылы біздің елімізде де машиналық аударманың дамып келе жатқанын көруге болады.

Сөздіктің электрондық адресі: <http://www.rubo.ru/sozdik.kz>.

Кез келген машиналық аударманы жасайтын электрондық аудармашыларда аударма белгілі бір жүйеленген алгоритм бойынша жүзеге асады.

Аталған мәселелер шешімі, теоретик-лексикографтар мен практиктердің өзара тығыз ынтымақтастығы арқасында шешімін табуы мүмкін, ал компьютерлік құралдар бірқалыпты лексикографиялық қиын жұмыстарды жеңілдетеді.

Электрондық сөздіктің жасалуына бағытталған компьютерлік лексикография – компьютерлік лингвистика бағытына аса қажет-

ті бағыт. Себебі осы бағытта жасалған сөздік өзінің жан-жақтылығымен, мультимедиялығымен, жаңа технологиялар шешімінің интеграциясымен, ақпараттың өзектілігімен ерекшеленеді және тұтынушының барлық сұраныстарына жауап береді.

### Семинар тапсырмалары

1. Электрондық сөздік дегеніміз не?
2. Сөздік құру қағидаларын атаңыз.
3. Дерекқор дегеніміз не? Дерекқордың анықтамасын келтіріңіз. Дерекқорды құрудағы негізгі әдістерді атаңыз.
4. *АВВУУ* Lingvo және *Multitran* электронды аударма сөздіктерінің ерекшеліктерін сипаттаңыз. Олардың онлайн аударма сөздіктерден (*google*, *yandex* және т.б.) айырмашылығы неде?
5. Әртүрлі электронды сөздіктерді сипаттаңыз: ерекшеліктері, бір-бірінен айырмашылығы, қолданысқа тиімділігі, кемшіліктері және болашағы.
6. Бір сөзді (қазақ, орыс, ағылшын тілдерінің) әртүрлі тілдерге әртүрлі сөздіктердің көмегімен аударып көріңіз де пікіріңізді айтыңыз.

### Әдебиеттер:

#### *Негізгі:*

1. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
1. Филиппович Ю., Чернышева М. Историческая компьютерная лексикография – terra incognita в компьютерном мире // Компьютерра. 1999. № 45. <http://offline.computerra.ru/1999/323/3379> (қаралған күн: 13.05.2011).
2. Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практ. руководство для студ., аспирантов, препод.-филологов. – М.: Флинта: Наука, 2007.
3. Селегей В. Электронные словари и компьютерная лексикография // Ассоциация лексикографов Lingvo. [www.lingvoda.ru/transforum/articles/selegey\\_a1.asp](http://www.lingvoda.ru/transforum/articles/selegey_a1.asp) (қаралған күн: 28.02.2012).

#### *Қосымша:*

1. Захаров В.П. Информационно-поисковые системы: учеб.-метод.пособие. – СПб., 2005. – С. 3-18.
2. Зубов А. И., Зубова И.И. Информационные технологии в лингвистике: учеб.пособие. – М.: Академия, 2004. – С.183-185.



## 12-дәріс. Компьютерлік терминография

**Қарастырылатын мәселелер:** *компьютерлік терминография түсінігі; термин – терминографияның негізгі объектісі; терминологиялық деректер қоры.*

**Сөздер мен сөз тіркестері:** *термин, терминография, компьютерлік терминография, терминологиялық дерекқор, терминологиялық сөздік, терминологиялық стандарт.*

Компьютерлік лексикография мен қолданбалы лингвистиканың басты бағыттарының бірі электронды терминологиялық сөздік және деректер қоры болып табылады. Арнайы терминологиялық сөздіктің құрылысымен терминография айналысады. Компьютерлік терминография дегеніміз – электронды терминологиялық сөздік жасау ғылымы.

Компьютерлік терминографияның қағидалары компьютерлік лексикографияның принциптерімен сәйкес келеді. Олардың айырмашылығы тек сөздіктің сипатындағы негізгі нысанмен ғана байланысты. Лексикографияда бұл – қарапайым сөз немесе басқа да тілдік бірліктер (морфемдер, сөз тіркестері, сөйлемдер және т.б.) болса, ал терминографияда негізгі нысан – термин.

Термин – ғылыми ұғымға айқын анықтама беретін, оның мағыналық шегін дәл көрсететін сөздер. Әдетте тілдегі қандай сөз болсын көп мағыналы болып келеді де оның мағыналық шегі айқын болмай, өзгермелі болады. Ал ғылыми ой-пікірді дәл білдіру үшін сөздің мағынасы тұрақты, айқын болу қажет. Сондықтан сөздің мағыналық шегін дәл белгілеп, сөзді сол нақтылы бір мағынада ғана алып қолдану арқылы жасалады. Мысалы, интернет сөзі бәрімізге таныс болғанымен, біз тек, қосу, өшіру, әлеуметтік желілерге кіру, тарифтер, провайдер және т.б. жалпы ұғымдармен таныспыз. Ал компьютер маманы үшін бұл сөз – компьютерлік серверлердің бүкіләлемдік желісі.

Интернетке қосылу мүмкіндігі болған жағдайда білім беру мекемелері, мемлекеттік ұйымдар, коммерциялық кәсіпорындар және жеке адамдар сияқты миллиондаған қайнар көзінен ақпарат алуға болады [1, 89-б.].

Термин ұғымы – белгілі бір саладағы немесе бағыттағы арнайы сөздер. Сонымен қатар терминдердің құрылымы терминдік жүйеге байланысты анықталады.

Қазір бүкіл әлемде 20 000-нан астам терминдік жүйені ұйымдастырудың стандарттары бар; одан өзге әртүрлі деңгейлерде: халықаралық, мемлекеттік және жекелеген компания мен фирмаларда терминологиялық стандарттар өмір сүреді [1, 90-б.].

Заманауи компьютерлік технологиялар әр мамандық саласы бойынша терминдерді өңдеп және сақтау мүмкіндігіне ие. Мұндай терминдер жиынтығы терминологиялық деректер қоры (ТДҚ) деп аталады. Деректер қорындағы тілдерге қатысты аударма (көптілді) және ақпараттық-нормативті (біртілді) терминологиялық деректер қоры болып ажыратылады. Көлемді ТДҚ Ғылыми-зерттеу институттарында, Халықаралық стандарттау ұйымдарында сақталады.

Терминологиялық сөздіктерге әр білім саласы бойынша Ғылыми-техникалық терминологиялық комитетінің жүйелі түрде шығарып тұратын ұсынылған 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.

### **Семинар тапсырмалары**

1. Арнаулы ғылым-техникалық терминологиялық сөздіктерді құрастырудың алға қойған мақсатына қарай, олардың әр мемлекетке тән терминографиялық дәстүрін сипаттаңыз. Бұл салыстырудан қандай нәтиже шығаруға болады?
2. Арнайы сөздіктерге қойылған міндет.
3. Өздеріңе таныс терминологиялық сөздіктерге анықтама беріңіздер.
4. Терминологиялық сөздіктің мақала аймағына не кіреді? Осындай аймақтардың бірін сипаттап беріңіздер.

### **Әдебиеттер:**

*Негізгі:*

1. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.

2. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб.пособие. М.: Академия, 2004. – С. 163-164.

*Қосымша:*

1. Захаров В.П. Информационно-поисковые системы: учеб.-метод.пособие. СПб., 2005. – С. 3-18.
2. Марчук Ю.Н. Компьютерная лингвистика: учеб.пособие. –М.: АСТ: Восток-Запад, 2007. – С. 190-195.

### **13-дәріс. Тілдерді компьютерлік оқыту**

**Қарастырылатын мәселелер:** *білім беру саласының ақпараттандырылуы және тілді үйретудегі өзгерістер; тілдерді компьютермен оқыту ұғымы; оқытудың электрондық құралдарының классификациясы; қашықтықтан оқыту; тілдерді оқытудағы желілік ресурстар.*

**Сөздер мен сөз тіркестері:** *шет тілдерін компьютер арқылы үйрету, бихевиорикалық кезең, когнитивті-зияткерлік кезең, компьютер-оқытушының көмекшісі, компьютер-оқытушы, компьютер-білім қоры, электрондық оқыту құралдары, компьютерлік оқулықтар, сынаушы бағдарламалар, жаттықтырушы бағдарламалар, оқу ойындары, компьютерлік анықтағыштар мен энциклопедиялар, қашықтықтан оқыту.*

Тіл білімінде компьютер қолданылатын маңызды қолданбалы салалардың бірі – тілдерді компьютермен оқыту (*Computer Assisted Language Learning, CALL*). Жалпы заманауи білім берудің негізгі ерекшеліктері болып ақпараттандыру мен компьютерлендіру саналады. Себебі білім беруде заманауи ақпараттық-коммуникациялық технологиялардың қолданылуы оқыту үрдісін тиімдірек етеді, үйренушілердің қызығушылығын арттырып, адам еңбегінің шығынын азайтады. Сонымен қатар компьютерлердің қолданылуы білім беру саласының қызметтік және жекеге бағытталған ерекшеліктері секілді өзге заманауи беталыстарына сәйкес келеді. Оған қоса шет тілдерін (ШТ) компьютер арқылы үйретуге баса назар аударылады, өйткені тіл үйрету үдерісі адам санасының барлық мүмдіктерінің пайдаланылуын қамтиды [1, 38-б.].

Білім берудегі компьютердің қолданылу мәселесі 1950 жылдардан, яғни компьютерлердің өндіріліп шығарылуынан бастап қарастырылып келеді [2, 3-б.]. Жарты ғасырлық уақыт аралығында компьютермен оқытуда түрлі елеулі өзгерістер болған. Бұл өзгерістердің өлшемі ретінде оқытудың басымырақ түскен тәсілдері мен компьютер техникасының даму деңгейі алынған. Компьютермен оқыту құралдарының дамуында екі маңызды кезең бар:

*Бихевиорикалық:* бұл кезеңдегі оқыту бағдарламасы «ынта-нәтиже» («стимул-реакция») формуласы бойынша құрылған, білім алушы *пассив* рөл атқарады, ал бағдарламалар жаттықтырушылар ретінде көрінеді.

*Когнитивті-зияткерлік:* бағдарламалар білім алушыға бағытталған, оған деңгей және әрекет түрін таңдау еркіндігін бере отырып, үйренушінің танымдық қызметтерін жандандырады [3, 111-б.; 1, 38 –39-б.].

Қазіргі таңда шет тілдерін компьютер арқылы үйрету бөлек білім саласы және тілдерді меңгеру мен үйретуге бағытталған өзіндік әдістемесі бар, бағдарламалық құралдар, мақсаты мен міндеттері қалыптасқан практикалық әрекеттердің жиынтығы ретінде көрініс табады [4, 1-б.]. Компьютерлерді қолдану мүмкіндіктері дәстүрлі бағдарлама-жаттығулардан бастап заманауи виртуальды оқыту құралдарына, мультимедиялық бағдарламалар мен қарым-қатынастың түрлі формаларының қолданыс табылуына, ғаламтордағы ақпараттың сақталуы, соның ішінде электронды пошта, корпустар мен конкорданстар, кәсіпшіліктер және т.б. беріледі.

Тілдерді компьютермен оқытудың ерекшелігі деп оқытудың негізі болатын белгілі бір теориялық әдісті (бихевиориялық, коммуникативті, когнитивті-зияткерлік және т.б.) танимыз. Сонымен қоса ол пәнаралық сипатқа ие: бүгінгі таңда тілдерді компьютермен оқыту мәселелері психологтардың, бағдарламашылардың, вебдизайнерлердің және тілшілердің бірлескен еңбектерімен шешіледі. Аталмыш пәнаралық бірлестікте компьютермен оқытудың жаңа өзгеше жолдары да шыға бастайды. Соның ішінде жасанды ақыл аясының әдістері мен ойлары тілдерді компьютермен оқытудағы жаңа бағыт – ICALL-ды (*Intelligent Computer Assisted Language Learning*) тудырды [1, 5; 58].

Тілдерді оқыту барысында компьютерлер әртүрлі жолдармен қолданылуы мүмкін:

– компьютер-оқытушының көмекшісі (дәстүрлі сабақтың жеке кезеңдерінде оқытушының компьютерді қолдануы);

– компьютер-оқытушы (тапсырылған қатаң жоспар бойынша тұтас оқу курсы жеке оқыту);

– компьютер-білім қоры және оны «бағалаушы» (қашықтықта тілдерді үйрену барысындағы жеке және топтық оқыту, осы тұста білім алушы компьютерді қажетті ақпарат көзі және алынған білімді «бағалаушы» ретінде таниды) [3, 138-139-б.].

Көріп отырғанымыздай, әр компьютерлік оқытудың өзіндік мақсаттық топтары мен шарттары бар. Бұл курсты меңгеруде оқытушыға көмек ретінде компьютердің қолданылуы міндетті болса, компьютерлердің оқытушы ретінде немесе білім көзі мен «бағалаушы» ретінде келуі білім алуда тұлғалық қажеттіліктерді қанағаттандырады.

Барлық жағдайларда сан алуан электрондық оқыту құралдары қолданылады [1,18-31; 6, 190-191; 7]:

- компьютерлік оқулықтар;
- сынаушы бағдарламалар;
- жаттықтырушы бағдарламалар;
- оқу ойындары;
- компьютерлік анықтағыштар мен энциклопедиялар және т.б.

*Компьютерлік оқулықтар* – оқу курсы немесе оның үлкен тарауын өз бетінше меңгеруге көмектесетін бағдарламалық-әдістемелік кешен. Ол қарапайым оқулық пен анықтағышты, тапсырмалар жинағы мен зертханалық жұмыстардың жиынтығы бола тұрып, оқытудың дәстүрлі түрлеріне балама емес, қосымша ретінде келеді.

*Сынаушы бағдарламалар* – қолданушыға сұрақ және жауаптың бірнеше нұсқалары берілетін арнайы компьютерлік бағдарлама. Аталмыш бағдарламаның басты мақсаты – қолданушының білім деңгейін тексеру. Кейбір жеңіл тесттер тіркелген стандартты сұрақтары мен қабылданған жауаптарды бағалайтын тұрақты жүйеге ие.

*Жаттықтырушы бағдарламалар* – белгілі бір коммуникативті әрекеттердің сол әрекеттердің көп мөлшерде қайталануы арқылы жасалуының автоматты дағдыларын қалыптастыратын бағ-

дарлама. Бағдарламалардың бұл түріне мысал ретінде соқыр басу әдісімен мәтінді клавиатурада жылдам басуды, етістіктің дұрыс түрлерін қолдану жаттығуларын және басқаларын айта аламыз.

*Оқу ойындары* – ойын және оқыту қызметтерін атқаратын компьютер бағдарламалары. Бұл жағдайда білім алушылар жалған рөлдерді ойнай отырып, білім алу және оны пайдалану үдерісіне белсенді араласады [Мәселен, төменгі Саксония неміс федерал жерлерінің ландтағын интерактивті жолмен игеретін оқу ойыны шыққан. Мұндай ойынның әр қатысушысы ландтағын виртуальды экскурсанты болып, осы заңды органның белгілі бөліктері туралы ақпаратты алғаннан кейін тесттік сынақтардан ойдағыдай өтіп, ғимараттың келесі қабатына көтеріле алады].

Оқу ойындарын түрлі мақсатта қолдануға болады:

– білім алушыларды жаңа білім қорын игеруге талпындыру үшін;

– ойын-сауық, демалыс, сабақ үстінде шаршағаннан айығу үшін;

– білім алушылардың қызығушылығын жандандыру үшін;

– өз бетінше оқуды дамыту үшін;

– жаттығу секілді үйренушілердің шеберлігін дамыту үшін.

*Компьютерлік анықтағыштар мен энциклопедиялар* – тек оқу материалын көрсетуге арналған бағдарламалар. Міндетті түрде автоматты іздеуді қажет ететін үлкен көлемдегі ақпараттарды қамтиды. Сонымен қатар аталмыш электрондық ресурстың ерекшелігі – қамтитын ақпараттың мультимедиялық сипаты және гипермәтінді ұйымы.

Оқытушы компьютерлік ресурстар дайын күйінде келуі мүмкін немесе оқытушылар дайын компьютерлік жаттығулардың көмегімен іске асыруы да ықтимал. Бұл ресурстардың алғашқы түрінің мысалы ретінде келесі мультимедиялық бағдарламалар келеді:

*Профессор Хиггинстің «Акцентсіз ағылшын тілі: ағылшын фонетикасы және грамматикасы»* бойынша мультимедиялық оқу құралы ([www.istrasoft.ru/higgins/htm](http://www.istrasoft.ru/higgins/htm));

*Bridge to English*: ересектерге арналған ағылшын лексикасы және грамматикасын оқыту бағдарламасы ([www.intense.ru](http://www.intense.ru));

Кроссвордтар, сөздері түсіп қалған сөйлемдер, сөйлемдері араласқан мәтіндер және өзге жаттығуларды құруға мүмкіндік беретін бағдарламалардың бірі – *Hot Potatoes* (<http://hotpot.uvic.ca>).

Түрлі электрондық оқу ресурстары заманауи қашықтықта оқытудың негізін құрайды. Қашықтықтан оқыту – өз бетінше білім алу қағидасына негізделген оқу үдерісінің ұйымдастырылуының формасы. Оқу материалының телекоммуникациялық жеткізілуі және оқу барысы, оны бағалау кезіндегі білім алушылар мен оқытушылардың интерактивті өзара арақатынасы ұстанымына негізделеді [3, 142 –143-б.].

Қашықтықтан оқыту келесі әрекеттерді қамтиды: білім алушымен қайта байланыс орнатуға мүмкіндік беретін тіркелу (курсқа жазылу), білім алушыға әртүрлі типтегі оқу материалдарының (мәтін, көрнекіліктер, видео, тапсырмалар және т.б.) берілуі және білім деңгейін анықтайтын тест тапсырмаларының орындалуы. Сонымен қоса білім алушы курс оқытушысынан чатта немесе электрондық пошта арқылы кеңес алуға мүмкіндігі бар.

Жеке тақырыптардың бөлінуі, теориялық материалдың таңдалуы, әр тақырыптың тапсырмалары мен жаттығулары, тест тапсырмаларының ыңғайлы жүйесі секілді ұтымды құрылысы қашықтықтан оқытуды ұйымдастыруда маңызды рөл атқарады.

Дидактикалық мақсаттағы түрлі веб-ресурстардың (электронды хаттар, веблогтар, кастішіліктер (подкасты), бірлескен вики-жобалар) қолданысқа ие болуы тілдерді компьютермен заманауи оқытудың ерекше бөлігі болып табылады. Жалпы *веб-ресурсты* ғаламторда енгізілген, веб-парақтары арқылы қолжетімді алуан түрлі ақпаратты (вербальды, графикалық, кестелі, дыбыстық, видеофайлдар, анимация және компьютерлік бағдарламалар) қамтитын электрондық құжат ретінде қабылдауға болады [2, 340-б.].

Веб-ресурстың қолжетімділік, ауқымдылық, жаһандық және бірдей келуі секілді ерекшеліктері тілдерді үйрену барысында оқу материалының ыңғайлы дереккөзі болатыны сөзсіз. Әйтсе де веб-ресурстар алуан түрлі болып та келуі мүмкін, ал олардың саны ғаламторда күн санап көбеюде. Мұндай жағдайда тілдерді үйрету мақсатында веб-ресурстарды қолданатын болсақ, ғаламтордағы қажетті оқу ақпараттарын дұрыс іздеуге бағытталған тапсырмалар және таңдай білу ұстанымы алғы қатарға шығады. Аталмыш тапсырмалардың үлгісі ретінде веб-квесттерді айта аламыз. Веб-квесті ғаламтор желісін қолдануда оқушылардың жобалық қызметін ұйымдастыратын жоспар ретінде түсіндіріледі [8, 97]. Шет тілдерін үйрету кезінде қолданылатын веб-квесттер-

дің тақырыптары алуан түрлі болу мүмкін: *My career* (Менің мансабым) [9, 98-101], *Extreme Sports* (экстремальды спорт), Джордж Гордон Байрон.

Веб-квест – Веб 2.0 (әлеуметтік желі) атауын иеленген веб-ресурстың түрі. Веб 2.0-ге адамдарға ақпаратты алумен бірге оны жасауға әрі бірігіп жасауға мүмкіндік беретін ғаламтордың әлеуметтік сервистері мен қызметтері жатады [8, 26-31].

Қорыта айтқанда, тілдерді компьютермен оқыту – заманауи лингводидактикадағы болашағы зор бағыт. Осы орайда айта кететін жайт, компьютерлік оқу ресурстарына оқытушының орнын басушы құрал ретінде қарамау керек. Керісінше, оларды күнделікті жұмыстың орындалуы мен ұйымдастырылуының дәстүрлі жолының кеңеюі, жаттығулар көмегімен білім алушылардың дағдыларын дамыту, студенттердің белсенділігін арттыру және өз бетінше оқуға мүмкіндіктерді қалыптастырудың жолы деп білу қажет.

### Семинар тапсырмалары

1. Тілдерді компьютермен оқытудағы бихевиорикалық және когнитивті-зияткерлік тәсілдерді сипаттаңыз.
2. Компьютерлік оқу ресурстарының артықшылықтары мен кемшіліктері?
3. Мультимедиялық оқытушы бағдарламалардың кезеңдерін баяндаңыз.
4. Мультимедиялық оқытушы бағдарламаларының классификациясының ерекшеліктерін атаңыз.
5. Қашықтықта оқытудың артықшылықтары мен кемшіліктері.
6. Оқу порталдары, электронды кітапхана, электронды нұсқадағы журналдар – веб-ресурстардың осы түрлерін қысқаша сипаттап беріңіз.

### Әдебиеттер

*Негізгі:*

1. Потапова Р.К. Новые информационные технологии и лингвистика: учеб. пособие. 2-е изд. – М.: Едиториал УРСС, 2004.
2. Розина И.Н. Педагогическая компьютерно-опосредованная коммуникация: теория и практика. – М.: Логос, 2005.
3. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. М.: Академия, 2004.
4. Levy M. CALL: context and conceptualization. Oxford: Oxford University Press, 1997.
5. Русский ассоциативный словарь / Ю.Н. Караулов, Ю.А. Сорокин, Е.Ф. Тарасов. Кн. 5: Прямой словарь: от стимула к реакции. – М.: ИРЯ РАН, 1998.



6. Villiger C. Lernsoftware // *Angewandte Linguistik: Ein Lehrbuch* / Hrsg. Von K. Knapp, G. Antos, M. Becker-Mrotzek u.a. Tübingen; Basel: Francke Verlag, 2004. – P. 187-206.
7. Warschauer M. *Computer Assisted Language Learning: an Introduction // Multimedia language teaching* / ed. by S. Fotos. Tokyo: Logos International, 1996. – P. 3-20.
8. Сысоев П.В., Евстигнеев М.Н. Методика обучения иностранному языку с использованием новых информационно-коммуникационных Интернет технологий: учеб.-метод. пособие. – М.: Глосса-Пресс; Ростов н/Д: Феникс, 2009.
9. Воробьева Е.И. Информатизация иноязычного образования: основные направления и перспективы. – Архангельск: Поморский университет, 2011.

*Қосымша:*

1. Зубов А. И., Зубова И.И. Информационные технологии в лингвистике: учеб.пособие. – М.: Академия, 2004. – С.110-114,138-145.
2. Потапова Р.К. Новые информационные технологии и лингвистика. Изд.2. – М.: Эдиториал УРСС, 2004. – С.18-42,117-118.
3. Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практ.руководство. – М., 2007. – С.10-30, 64-100.

## 14-дәріс. Корпустық лингвистика

**Қарастырылатын мәселелер.** *корпустық лингвистика қолданбалы лингвистиканың бір бөлігі; корпус ұғымы, белгілеу (разметка), корпус түрлері; корпус талаптары.*

**Сөздер мен сөз тіркестері:** *корпустық лингвистика, мәтіндер корпусы, мәтін лингвистикасы, лингвистикалық корпус, мәтіндерді жинау, репрезентативтілік толықтылық, жеткілікті көлем, үнемділік, материал құрылымы, компьютерлік қолдау, корпус тар классификациясы, сақталу формасы, мәтіндер тілі, «параллельдік», стиль.*

Лингвистиканың маңызды міндеттерінің бірі лингвистикалық зерттеулер үшін нақты материалдар дереккөздерін сақтау және жинау болып табылады. Қазіргі таңда бұл міндетті шешу мақсатында электронды түрде сақтауға қолайлы болып табылатын саналуан қолданыстық бағытта көлемді мәтіндерді жинау қол-

данылады. Компьютер мен арнайы телекоммуникациялық желілерді тарту тек электронды түрде көлемді мәтіндерді жинап қана қоймай, оларды өңдеу мен іздеуге де мүмкіндік береді. Корпустар мен мәтіндерді электронды түрде жинақтау қазіргі лингвистика ғылымы үшін маңызды болып табылғандықтан, бұл электронды мәтіндерді жинақтау арнайы қолданбалы лингвистика – корпусты лингвистиканың зерттеу нысанына айналды.

Корпустық лингвистика – лингвистикалық корпустарды компьютердің көмегімен қолдану және жалпы құру қағидаларын құрастырумен айналысатын қолданбалы лингвистиканың бөлігі [1, 3-б.].

Бұл анықтамаға сай, корпустық лингвистика екі аспектіні қамтиды:

1) автоматты құралдар қолданатын мәтіндер корпусын құрастыру;

2) әртүрлі типтегі корпустар базасында тілдің әртүрлі деңгейлерінде эксперименталды зерттеулер тәсілін құрастыру [2].

Қазіргі зерттеуші-лингвистер өздерінің жеке корпустарын жасай алады, содан соң сол базада өзіндік қажетті зерттеулерін жүргізе алады, сонымен қатар басқа зерттеушілер және олардың ұжымы жасаған жалпыға ортақ корпустарды қолдана алады.

Корпустар ғылыми зерттеулер жүргізгеннен басқа [3, 166-167; 4, 60]:

1) *лексикографияда* көпмағыналы сөздерді анықтау және сөздік жасауда;

2) *грамматикада* морфемалардың жиілігін анықтау, сөз тіркестерінің және сөйлемдердің типін анықтауда;

3) мәтін *лингвистикасында* мәтін типтерінің дифференциясын, азатжолдың ішкі байланыстары мен азатжол арасын анықтауда және т.б.;

4) *автоматты аудармада* мәтіндердің параллель мәтіндердегі аударма баламаларын, бірнеше баламаға ие сөздерді іздеуде;

5) *оқыту мақсатында* шығарма үзінділерін, оқу орындары үшін үлгілер, оқу құралдарын жасау, дәйексөз алуда және т.б.

6) *тестілеу* бағдарламасында автоматты анализ бен сөйлеу синтезі және тағы басқа жағдайлар үшін қолданылады.

Корпустық лингвистиканың орталық ұғымы – *лингвистикалық корпус* – іздеу жүйесі қамтылған түрлі лингвистикалық параметрлермен белгіленген арнайы жинақталған мәтіндер жиынты-

ғы ретінде анықталады. Корпусты келесі үлгіде қысқаша былай көрсетуге болады:

Корпус = мәтіндер + және олардың белгілері (разметка)

Ал жалпы кең мағынада корпус дегеніміз – кез келген мәтіндер жиынтығы. Бұл жағдайда мәтіндердің белгіленген және белгіленбеген корпустары жіктеледі. Белгіленбеген корпустар санына мәтіндердің электронды жиынтықтарын қарастырамыз: виртуалды кітапханалар, біршама зерттеулерге жеткілікті болып табылатын мерзімді басылымдардың электронды мұрағаттары жатады. Алайда белгіленбеген мәтіндер жиынтығымен жұмыс жасау зерттеуші үшін релевантсыз болып табылады, бұндай дереккөзбен жұмыс жасау қиынға соғады. Осыған байланысты, корпустық лингвистиканың пәніне мәтіндердің белгіленген корпустары жатады.

Корпусты құрудың бірінші кезеңіне мәтіндерді жинақтау жатады. Бұл кезеңде мәтіндердің қай жанр, қандай стильде, сонымен қатар хронологиялық кезеңі қарастырылуы қажет. Корpusқа мәтіндерді жинақтауда корпус құруда біршама қағидаларға сүйену қажет [5, 118 – 119-б., 2]:

1) *репрезентативтілік* (корпустағы құбылыс жиілігі табиғи тілдегі жиілікпен сай болуы қажет);

2) *толықтылық* (полнота) (кез келген құбылыс корпуста қамтылуы тиіс);

3) *жеткілікті көлем* (егер алғашқы корпустар миллион сөзден тұрса, ал қазіргі таңдағы корпустар миллиардтан асып кетеді);

4) *экономичность/үнемділік* (мәтіндер корпусы проблемалық саланы зерттеуші үшін үнемді болуы керек);

5) *материал құрылымы*;

6) *компьютерлік қолдау* (деректерді өңдеу белгілі бір компьютерлік бағдарламалармен жасакталуы керек).

Белгілеу – корпус құрастырудың маңызды кезеңі. Белгілеу (разметка) (англ. *tagging, annotation*) – мәтіндер мен оның компоненттеріне арнайы белгіні салу. Бұл белгілер автор және мәтін туралы мағлұматты қамтитын сыртқы (экстралингвистикалық) немесе ішкі болады. Сыртқы белгілер автор жайлы мағлұмат, мәтіннің атауы, жылы және басылған жері, жанры, тематикасы туралы ақпаратты қамтиды. Автор жайлы мағлұмат тек атын ғана емес, сонымен қатар жасын, жынысы және тағы басқа мәліметтерді

қамтуы мүмкін. Бұл ақпаратты кодтау метабелгілеу атауына ие болады. Құрылымдық белгілер әрбір бірліктің жағдайы туралы ақпаратты қамтиды (бөлім, азатжол, сөйлем, сөзжасам), ал лингвистикалық белгілер лексикалық, грамматикалық және мәтіннің өзге элементтерін сипаттайды [6, 6-б.].

Лингвистикалық сипаттамалар деңгейлерге сай әртүрлі морфологиялық (сөз табы мен категорияларын анықтау), синтаксистік (синтаксистік байланыстарды анықтау), семантикалық (сөз мағынасын сипаттайтын категориялар), просодикалық (екпін және интонация), дискурстық (кідіріс, қайталау) және басқа да белгілеу (разметка) түрлері [6, 6 – 7-б.].

Корпустағы мәтіндердің түрлеріне және олардың белгілеріне (разметка), сондай-ақ басқа да факторларға байланысты келесі корпустың түрлері ажыратылады (2-кесте).

2-кесте

### Корпустар классификациясы

№	Белгі	Корпус түрлері
1	Сақталу формасы	Дыбыстық, жазу, аралас
2	Мәтіндер тілі	Қазақ, орыс, ағылшын және т.б.
3	«Параллельдік»	Біргілді, көптілді, көптілді
4	Стиль	Әдеби, диалектілік, ауызекі Публицистикалық, терминологиялық, аралас
5	Қолжетімділік (доступ) тәсілі	Еркін, коммерциялық, жабық
6	Белгілеу (разметка)	Белгіленген, белгіленбеген
7	Белгілеу (разметка) сипаты	Морфологиялық, синтаксистік, семантикалық, просодикалық және т.б.

Корпус түрлерінің ең маңыздысы – әртүрлі ұлттық тілдерге арналған әмбебап ұлттық корпус. Әмбебап ұлттық корпустың құрылуы мен кеңеюі корпустық лингвистиканың негізгі міндеттерінің бірі болып табылады. Қазіргі таңда жалпы қолданысқа ие бірнеше корпустар бар [3, 169-б.; 2]:

- Британдық ұлттық корпус (BNC);
- Орыс тілінің ұлттық корпусы;

- Чеш ұлттық корпусы;
- Браун корпусы және т.б.

Белгіленген мәтіндер жинағы жинақталған корпустар қолданбалы мүмкіндіктерге ие болып табылады.

### Семинар тапсырмалары

1. Корпус дегеніміз не?
2. Корпус пен электрондық сөздіктің қандай айырмашылықтары бар? Мысалмен дәйектеңіз.
3. Қандай корпусты білесіздер?
4. Қазіргі таңда қазақ тілінің ұлттық корпусын жасаудың алғышарттары қандай?
5. Берілген корпустың бірін таңдап, оған сипаттама беріңіз (британдық ұлттық корпус ([www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk)), американдық ұлттық корпус ([www.americannationalcorpus.org](http://www.americannationalcorpus.org)); ағылшын тілінің банкі (Bank of English) ([www.collins.co.uk/Corpus/CorpusSearch.aspx](http://www.collins.co.uk/Corpus/CorpusSearch.aspx)); орыс тілінің ұлттық корпусы ([www.ruscorpora.ru](http://www.ruscorpora.ru))).

### Әдебиеттер:

#### Негізгі:

1. Захаров В.П. Корпусная лингвистика: учеб.-метод. пособие. СПб.: СПбГУ, 2005.
2. Толдова С.Ю., Архипов А.В., Логинова Е.А., Попова Д.П. Корпусная лингвистика // Фонд знаний «Ломоносов». – М., 2011. [www.lomonosov-fund.ru/enc/ru/encyclopedia:01210:article](http://www.lomonosov-fund.ru/enc/ru/encyclopedia:01210:article) (қаралған күн: 28.02.2012).
3. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
4. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации: учеб.-метод. пособие. – М., 2009.
5. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.

#### Қосымша:

1. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007. – С. 112-137.
2. Захаров В.П. Корпусная лингвистика: учебно-метод. пособие. СПб.: СПбГУ, 2005. – С. 3-14.
3. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации: учебно-метод. пособие. – М., 2009. – С. 60-76.

## 15-дәріс. Ақпараттық-іздеу жүйелері

**Қарастырылатын мәселелер:** *ақпараттық-іздеу жүйесі ұғымы; галамтордағы іздеу құралдарының түрлері; галамтор іздеу жүйесінің сипаты; ақпараттық-іздеу тілі.*

**Сөздер мен сөз тіркестері:** *ақпаратты іздеу, каталогтар, сілтемелер топтамасы, іздеу машиналары, электрондық пошта мекенжайларының мәліметтер қоры, веб-ресурстар каталогы, релеванттылық категориясы іздеудің толықтығы, іздеудің нақтылығы.*

Ақпарат көлемі күннен-күнге ұлғайып келетін қазіргі таңда мәліметтерді білім қоры ретінде пайдаланатын адам үшін бейімделу мәселесі орын алады. Сондықтан ақпараттар легінде «адаспас» үшін оларды таңдау, сүзу және бағалау техникасын қажет етеміз [1, 18-б.].

Ақпаратты сүзу және таңдаудың дәстүрлі тәсілдері төмендегідей:

- «жоғары» («жоғарыдан») іздеу (мазмұны бойынша);
- «төменнен» іздеу (әртүрлі сілтемелер көмегімен жүзеге асады);
- гипермәтіндік байланыстар арқылы іздеу (байланысқан сілтемелер);

– тұтас мәтінді қарау арқылы іздеу [2, 70-б.] .

Іздеудің соңғы түрі өзгелеріне қарағанда нақты, дегенмен көп еңбекті және уақытты қажет етеді.

Ізденістің ұйымдастырылуы бірнеше сатыны қамтиды. Олар:

– Іздеу жұмыстарына негіз болатын көптеген құжаттар (мәтіндер немесе олардың үзінділері);

– Қолданушының ақпараттық сұранысында көрініс табатын ақпаратқа деген коммуникативтік қажеттілік;

– Ақпараттық сұранысқа сәйкес келетін бастапқы деректің белгілі бөлігінде коммуникативтік қажеттіліктердің қанағаттандырылуы [3, 197-б.].

Мәтіндер немесе олардың бөліктері ретінде келген ақпараттың іздестірілуі мен сақталуы үшін арналған құжаттар мен ақпараттық технологиялардың реттеулі жиынтығы ақпараттық-іздеу жүйесі деп аталады (АІЖ) [4, 3-б.].

Адам күшін үнемдеу мақсатында 1950 жылдардан бастап АІЖ-ны автоматтандыру ісі қолға алынып келеді. Алғашқы АІЖ-

да құжаттардың мазмұнын сипаттау мен талдау (индексациялау) қолмен жүргізілді, ал бұл құжаттар бойынша іздеулер автоматты түрде іске асты [4, 8-б.].

Бүгінгі таңда компьютер техникасының дамуы мен жоғары жылдамдықтағы телекоммуникациялық желілердің жасалуы автоматтандыру саласында, Б. Гейтстің атақты «саусақ ұшындағы ақпарат» (*information at your fingertips*) формуласында көрсетілгендей, маңызды жетістіктерге қол жеткізілді [1, 16 бойынша]. Бұл пікірге сенер болсақ, ақпарат әрдайым адам «әмірінде» болады, тек оған жету үшін батырманы басу қажет екен.

Ендеше ғаламторда ақпаратты іздеуге арналған іздеу құралдарының сан түрлері қызмет етеді [2, 71-б.]:

- *Каталогтар (directories)*;
- *Сілтемелер топтамасы (bookmarks)*;
- *Іздеу машиналары (search engines)*;
- *Электрондық пошта мекенжайларының мәліметтер қоры және т.б.*

Іздеу құралдарының әрқайсысы өзіндік ерекшеліктерге ие. Егер адам іздеу мақсаты жөнінде толық мағлұматқа ие болмаса, онда веб-ресурстар каталогтарын қолданған жөн. Ал егер пайдаланушы өзіне қажетті ресурсты сипаттайтын тірек сөздерді білсе, іздеу машиналарын қолданған тиімді.

*Веб-ресурстар каталогы* – категориялардың иерархиялық құрылысы бойынша жіктелген ресурстардың үнемі толықтырылып және жаңаланып тұратын сілтемелер жүйесі. Каталогтың ең жоғарғы деңгейінде «ғылым», «бизнес» «сауық» секілді ортақ категориялар (айдарлар) беріледі. Төменгі деңгей айдарлары жекелік сипатқа ие [2, 71-б.]. Мәселен, «ғылым» айдары бірнеше категорияларға жіктеледі: «нақты ғылымдар», «табиғи ғылымдар» және гуманитарлық ғылым». Ал соңғы аталғаны философия, әлеуметтану, психология, педагогика және т.б. болып жіктеледі. Орыс тілді парақшалар каталогын мысалы, мына мекенжай бойынша [www.ru](http://www.ru) табуға болады.

*Сілтемелер топтамасы* – ғаламтордағы ақпаратты ұымдас-тырудың тағы бір әдісі. Аталмыш топтаманы арнайы мамандар белгілі бір тақырыпта жасайды, үнемі жаңаланып тұрады әрі қажет емес ақпаратты қамтымайды. Тіл білімінде ақпараттық технологияларды қолдануға байланысты сілтемелер топтамасының

жазба түрін біздің оқу құралымызда библиографиялық тізімнен кейін таба аласыз. Ағылшын тілін оқыту бойынша сілтемелер жинағынан үлгілерді С.В. Титова келтіреді [5, 27 – 28-б.].

*Іздеу машиналары (немесе іздеу жүйелері)* – пайдаланушының сұранысына сәйкес келетін мәтінді қамтитын веб-ресурстарды табуға мүмкіндік беретін арнайы веб-парақшалар. Іздеу машиналарының халықаралық каталогында ([www.searchenginecolossus.com](http://www.searchenginecolossus.com)) 232 елден 2300-ден аса іздеу жүйелері тіркелген. Осы каталогтың мәліметтері бойынша күн сайын 450 млн іздеу сұраныстары жасалады [2, 72-б.; 6, 364-б.].

Атақты іздеу машиналарының қатарында [2, 72 – 73-б.]:

- *AltaVista* ([www.altavista.com](http://www.altavista.com));
- *Excite* ([www.excite.com](http://www.excite.com));
- *Yahoo!* ([www.yahoo.com](http://www.yahoo.com));
- *AOL* (<http://search.aol.com>);
- *MSN* (<http://search.msn.com>);
- *Google* ([www.google.ru](http://www.google.ru));
- *Yandex* ([www.yandex.ru](http://www.yandex.ru));
- *Rambler* ([www.rambler.ru](http://www.rambler.ru));
- *Anopm* ([www.aport.ru](http://www.aport.ru)).

Іздеу жүйесінде іздеу жұмыстарының қалай жүретініне назар аударсақ, пайдаланушы өзінің іздеу сұранысын арнайы жолға енгізеді. Жалпы табиғи тілде жазылған сұраныс іздеу бағдарламасы арқылы *ақпараттық-іздеу (АІТ)* тіліне айналады. АІТ – сұранысты және АІЖ-да сақталатын құжаттар мазмұнын сипаттауға арналған шартты тіл [3, 201-б.]. Ақпараттық іздеу тілі өзіндік әліпбиін, лексикасын, грамматикасын, таңбалар жүйесін және оларды қолдану ережелерін қамтиды. Әрбір іздеу жүйесінің АІТ ерекшеліктерін, әсіресе оның «синтаксисі» туралы (іздеу жолына жазылатын тірек сөздер тіркесу ережелері) аталмыш жүйенің жекелеген қосымша парақтарынан табуға болады. Мәселен, *Яндекс*те осындай парақ «*Көмек-Қалай іздеу керек?*» деп аталады.

АІТ-да құжаттарды сипаттау үдерісі *индекстеу* деп аталады. Индекстеу нәтижесінде әр құжатқа оның шартты сипаты, яғни құжаттың іздестіру бейнесі тіркеледі. Аталмыш жолмен сұраныс та индекстеліп, оған сұраныстың іздестіру бейнесі немесе іздестіру бұйрығы тіркеледі. Ақпараттық іздестірудің жоспары негізін-



де іздестіру бұйрығымен сұраныстың іздестіру бейнесін салыстыру жатыр [3, 201-б.].

Құжаттың сұранысқа сәйкес келуі *релеванттылық* категориясымен өлшенеді. Осы тұста ақпараттық іздеу нәтижесінде мағынасы жағынан релевантты емес, дегенмен шартты түрде релевантты келетін көптеген құжаттар күйіндегі *ақпараттық айқай-шуды* алуға болады [3, 197 – 198].

Пайдаланушы іздестіру жүйесінің қосымшаларын қолдана отырып, өзінің сұранысын нақтыласа, ақпараттық шу да аз болу мүмкін. Мәселен, *Google*-да «*Ауқымды іздестіру*» парағын ашса, тұтас сөз тіркестерін (оны құраушы әр сөзді емес) іздеуге, шығарудың тілін, құжаттың жасалған күнін, сөз қолданылған құжаттың бөлігін, құжаттың форматын және т.б. тежеуге болады. Осындай әрекеттер арқылы қажетті ақпарат берілетін тізімнің басында орналасуы мүмкін.

Іздестіру нәтижелері екі түрлі көзқараспен, нақтырақ айтсақ, толықтығы және нақтылығы жағынан сипатталады. *Іздеудің толықтығы* (ағылш. *Recall*) деп ақпарат легінде қамтылған релевантты құжаттардың ортақ саны мен берілген релевантты құжаттардың мөлшері арасындағы қатынасты білдіретін өлшемді айтамыз. *Іздеудің нақтылығы деп* (ағылш. *Precision*) берілген релевантты құжаттардың мөлшерінің шығарылатын құжаттардың жалпы санына қатынасын көрсетеді [4, 8-б.].

Түрлі іздеу жүйесінің нәтижелерін салыстыру арқылы іздестірудің толықтығы мен нақтылығы туралы кешенді мәлімет алуға болады. Сұраныс тірек сөздерінің және олардың синтаксистік байланысын тура анықтаған жағдайда іздестірудің толықтығы мен нақтылығы бір болады, яғни іздеудің қажетті нәтижесін таңдауды жеңілдететін релевантты құжаттарды азайтады.

Сонымен, ақпараттар «саусақ ұшында ғана беріліп қоймайды», таңдаудың және іріктеудің күрделі жолдарымен беріледі. Аталмыш міндеттерді жүзеге асыруға ғаламтордың іздеу жүйесі секілді заманауи автоматты ақпараттық іздестіру жүйелері септігін тигізеді.

### Семинар тапсырмалары

1. Іздеу шартты және мағыналы релеванттылығы дегеніміз не? Бұл ұғымдардың айырмашылығы іздеу нәтижелерінде қалай көрінеді?

2. Ақпараттық іздеу жүйесінің екі маңызды түрі: құжаттық және фактографиялық жүйесін сипаттап беріңіз.
3. Ақпараттық іздеу жүйесінің автоматты және қолмен индекстеуден қандай айырмашылықтары бар? Жүйенің екі түріне мысал келтіріңіз.
4. Веб-ресурстың жалпы және арнайы каталогы дегеніміз не? Каталогтың екі түріне мысал келтіріңіз.

### Әдебиеттер:

#### Негізгі:

1. Больц Н. Азбука медиа / пер. с нем. Л. Ионина, А. Черных. – М.: Европа, 2011.
2. Башмаков И.А, Башмаков А.И. Интеллектуальные информационные системы. – М.: МГТУ им. Н.Э. Баумана, 2005.
3. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
4. Захаров В.П. Информационно-поисковые системы: учеб.-метод. пособие. СПб.: СПбГУ, 2005.
5. Титова С.В. Ресурсы и службы Интернета в преподавании иностранных языков. – М.: Изд-во «МГУ», 2003.
6. Розина И.Н. Педагогическая компьютерно-опосредованная коммуникация: теория и практика. – М.: Логос, 2005.

#### Қосымша:

1. Баранов А.Н. Введение в прикладную лингвистику: учеб.пособие. – М.: Эдиториал УРСС, 2001. – С.197-207.
2. Захаров В.П. Информационно-поисковые системы: учеб.-метод.пособие. СПб., 2005. – С. 3-18.

## 2-АРАЛЫҚ БАҚЫЛАУ БОЙЫНША ТЕСТ ТАПСЫРМАЛАРЫ

### 1. Машиналық аударманы басқаша қалай атауға болады?

- A) анықтамалық аударма
- B) автоматты аударма
- C) синтезді аударма
- D) графемалық аударма

### 2. Машиналық аударма дегеніміз не?

- A) бір табиғи тілдегі мәтінді басқа тілге арнайы компьютерлік бағдарламамен аудару үдерісі
- B) жұмыстың көп бөлігін, жартысынан астамын машина орындайтын аударма
- C) аударма жасауға жұмсалған еңбекті бөлу
- D) мәтін сөйлемінің грамматикалық құрылымын айқындау

**3. Аударма алгоритмі дегеніміз не?**

А) бір табиғи тілдегі мәтінді компьютер арқылы мәтін мазмұнының баламаларымен басқа тілге айналдыратын іс-әрекет, сонымен бірге осы іс-әрекеттің нәтижесі, қорытындысы

В) компьютер арқылы орындалатын, бір табиғи тілдегі мәтіндерді басқа тілге мәтін мазмұнының баламалары (эквивалент) бойынша қайта құратын (жаңадан жасау) іс-әрекет, сонымен қатар мұндай іс-әрекеттің қорытындысы, нәтижесі

С) бір мағыналық жүйе ретінде және аударма бағытында берілген тілдер жұбына (пара языков L1–L2) нақты бір тілден басқа тілге мәтін бойынша қатаң анықталған аударма сәйкестілігін табу үшін жүргізілетін іс-әрекет

Д) бір табиғи тілдегі әртүрлі мәтіндерді басқа тілге аудару үдерісін жартылай немесе толық компьютер арқылы жүзеге асыру

**4. Машиналық аударма жүйесінде талдау мен жинақтауды қамтамасыз ететін ең кең таралған формалды операциялар жүйелілігі, бірізділігі қанша кезеңнен тұрады?**

А) 3

В) 4

С) 2

Д) 5

**5. Машиналық аударма жүйелерін аударма барысында адамның қатысуына байланысты жіктеуді кім ұсынған?**

А) Л. Чайлдс

В) Ф.де Соссюр

С) В. Гумбольдт

Д) Г. Штейнталь

**6. НАМТ (Human-Assisted Machine Translation) дегеніміз не?**

А) толық автоматтандырылған машиналық аударма жүйелері

В) аударманы бағдарлама (компьютер) құралдарын қолдану арқылы адамның өзі аударатын жүйе

С) компьютерден аударылып шығатын сөз формалар мен тұтас сөйлемдерді компьютерден, машинадан шығатын тілде жинақтайтын жүйе

Д) адамның қатысуы арқылы жүзеге асатын машиналық аударма

**7. А.Н. Баранов машиналық аударманың компьютерлік әдістерін қандай топтарға бөледі?**

А) машиналық аударма жүйесі (МП)

- В) адам-машина аударма жүйесі (ЧМП), машиналық аударма жүйесі (МП) терминологиялық деректер қоры (ТБД)  
С) машиналық аударма жүйесі (МП), терминологиялық деректер қоры (ТБД)  
D) терминологиялық деректер қоры (ТБД)

**8. L-MASTER – 98 электрондық сөздігі қай қалада құрастырылған?**

- A) Париж  
B) Мадрид  
C) Киев  
D) Дели

**9. PROMT компаниясының қанша электрондық сөздігі бар?**

- A) 5  
B) 3  
C) 4  
D) 2

**10. Электрондық сөздік дегеніміз не?**

- A) бір табиғи тілдегі мәтінді басқа тілге арнайы компьютерлік бағдарламамен аудару үдерісі  
B) бір табиғи тілдегі әртүрлі мәтіндерді басқа тілге аудару үдерісін жартылай немесе толық компьютер арқылы жүзеге асыру  
C) машинаға енгізген мәтінді ешбір қосымша көмексіз басқа тілде шығаратын үдеріс  
D) қажетті сөздерді тез аударуға, сөздіктен іздеуге және аударма бағытын, мысалы, ағылшын-орыс тілдерінен, орыс-ағылшын тілдеріне өзгертуге мүмкіндік беретін арнайы сөздік мақалаларынан, сөздіктерден құралған компьютерлік деректер қоры

**11. Автоматтық сөздіктің қанша түрі бар?**

- A) 2  
B) 3  
C) 4  
D) 5

**12. ABBYY Lingvo электрондық сөздіктер жүйесінің ең соңғы нұсқасы қашан шықты?**

- A) 2007 ж.  
B) 2008 ж.  
C) 2006 ж.  
D) 2005 ж.

**4. Multitran электрондық сөздігі қандай болып келеді?**

- A) толық мәтіндік аударма жасай алмайды
- B) орыс, ағылшын, неміс, француз, испан, итальян, нидерланд, латыш, эстон және жапон тілдерінен аударма жасайтын электрондық сөздіктер жүйесі
- C) ағылшын тілінен орыс тіліне және керісінше аударарды
- D) жалпы лексикалық және арнайы мамандыққа арналған сөздіктер қоры кеңейтілген электрондық сөздіктер жүйесі

**5. Multitran электрондық сөздігінің бағдарламасын құрастырған кім?**

- A) А.Н. Баранов
- B) А. Рязанов
- C) А. Поминов
- D) Г. Штейнталь

**6. CD-ROM дискісіне “ЭТС” баспасы басып шығарған сөздік қалай аталады?**

- A) Polyglossum сөздігі
- B) Multitran сөздігі
- C) Context сөздігі
- D) Start Dict сөздігі

**7. Қазақша-орысша электрондық сөздік қалай аталады?**

- A) Context сөздігі
- B) Polyglossum сөздігі
- C) Sozdik
- D) Multitran сөздігі

**8. Термин дегеніміз не?**

- A) ғылыми ұғымға айқын анықтама беретін, оның мағыналық шегін дәл көрсететін сөздер
- B) машинаға енгізген мәтінді ешбір қосымша көмексіз басқа тілде шығаратын үдеріс
- C) жұмыстың көп бөлігін, жартысынан астамын машина орындайтын аударма
- D) дыбыстардың фото-бейнесі

**9. Әлемде терминдік жүйені ұйымдастырудың қанша стандарты бар?**

- A) 5480
- B) 30
- C) 20000-нан астам
- D) 2500-дей

**10. Белгілі бір ғылым саласына тән сөздер?**

- A) неологизмдер
- B) терминдер
- C) диалект сөздер
- D) историзмдер

**11. Компьютермен оқыту құралдарының дамуында қандай кезеңдерді ажыратады?**

- A) бихевиорикалық
- B) когнитивті-зияткерлік
- C) бихевиорикалық, когнитивті-зияткерлік
- D) реалды

**12. Сынаушы бағдарламалар дегеніміз не?**

- A) қолданушыға сұрақ және жауаптың бірнеше нұсқалары берілетін арнайы компьютерлік бағдарлама
- B) аударманы бағдарлама (компьютер) құралдарын қолдану арқылы адамның өзі аударатын жүйе
- C) мәтінде әрбір қолданысқа түскен сөздің бастапқы тұлғасын анықтау және осы сөздің морфологиялық сипаттарының жинағы
- D) кейбір зат немесе құбылысты білдіретін таңба

**13. Лингвистикалық корпустарды компьютердің көмегімен қолдану және жалпы құру қағидаларын құрастырумен айналысатын қолданбалы лингвистиканың бөлігі қалай аталады?**

- A) когнитивтік лингвистика
- B) корпусдық лингвистика
- C) қолданбалы лингвистика
- D) психолингвистика

**14. Корпусдық лингвистиканың орталық ұғымы?**

- A) таңба
- B) сөздік
- C) лингвистикалық корпус
- D) мәтін

**15. Корпусқа мәтіндерді жинақтауда, корпус құруда қанша қағидаға сүйену қажет?**

- A) 6
- B) 5
- C) 3
- D) 4

**16. Ақпараттық-іздеу жүйесін автоматтандыру ісі қай уақыттан бастап қолға алынып келеді?**

- A) 1950 ж.
- B) 1945 ж.
- C) 1956 ж.
- D) 1970 ж.

**17. «Саусақ ұшындағы ақпарат» формуласын кім ойлап тапқан?**

- A) А.Рязанов
- B) А.Поминов
- C) Б. Гейтс
- D) А.Н. Баранов

**18. Веб-ресурстар каталогы дегеніміз не?**

A) аударманы бағдарлама (компьютер) құралдарын қолдану арқылы адамның өзі аударатын жүйе

B) қолданушыға сұрақ және жауаптың бірнеше нұсқалары берілетін арнайы компьютерлік бағдарлама

C) категориялардың иерархиялық құрылысы бойынша жіктелген ресурстардың үнемі толықтырылып және жаңаланып тұратын сілтемелер жүйесі

D) қажетті сөздерді тез аударуға, сөздіктен іздеуге және аударма бағытын, мысалы, ағылшын-орыс тілдерінен, орыс-ағылшын тілдеріне өзгертуге мүмкіндік беретін арнайы сөздік мақалаларынан, сөздіктерден құралған компьютерлік деректер қоры

**19. Корпус құрастыруда ұстанатын қағидаларды көрсетіңіз:**

- A) репрезентативтілік, толықтылық,
- B) көлемдік, үнемділік
- C) деректерді өңдеу, толықтыру
- D) аталғандардың бәрі

**20. Электрондық кітапханаға тән белгіні табыңыз:**

- A) лингвостатистика
- B) лингвистикалық белгі
- C) толық мәтіндер
- D) тілдік материалдың репрезентативтілігі

**21. Мәтіндердің лингвистикалық корпусына тән емес белгіні көрсетіңіз:**

- A) лингвостатистика
- B) лингвистикалық белгі
- C) библиографиялық элементтер
- D) тілдік материалдың репрезентативтілігі

**22. Мәтін тіліне байланысты корпустар қалай жіктеледі?**

- A) орыс, қазақ, ағылшын және т.б.
- B) біртiлді, қостiлді
- C) әдеби, терминологиялық
- D) фольклорлық, драматургиялық

**22. Жанрлық белгісіне қарай корпустардың жіктелуін көрсетіңіз:**

- A) ауызша
- B) біртiлді
- C) терминологиялық
- D) публицистикалық

**23. Белгілеу сипатына жатпайтын қатарды табыңыз:**

- A) морфологиялық
- B) синтаксистік
- C) семантикалық
- D) фрагменттік

**24. Іздеу машиналары дегеніміз –**

- A) пайдаланушының сұранысына сәйкес келетін мәтінді қамтитын веб-ресурстарды табуға мүмкіндік беретін арнайы веб-парақшалар
- B) ғаламтордағы ақпаратты ұйымдастырудың тағы бір әдісі
- C) категориялардың иерархиялық құрылысы бойынша жіктелген ресурстардың үнемі толықтырылып және жаңаланып тұратын сілтемелер жүйесі
- D) берілген релевантты құжаттардың мөлшерінің шығарылатын құжаттардың жалпы санына қатынасы

**25. Ғаламторда ақпаратты іздеуге арналған іздеу құралдарын көрсетіңіз:**

- A) сілтемелер топтамасы
- B) іздеу машиналары
- C) электрондық пошта мекенжайларының мәліметтер қоры
- D) аталғандардың бәрі



## ӘДЕБИЕТТЕР

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011.
2. Алексеев В. Услышь меня, машина // Компьютерра. 1997. № 49. <http://offline.computerra.ru/1997/226/938> (қаралған күн: 28.02.2012).
3. Апресян Ю.Д. Идеи и методы современной структурной лингвистики. – М.: Просвещение, 1966.
4. Аханов К. Тіл білімінің негіздері. – Алматы, 2004. – 38 б.
5. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. – М.: ЛКИ, 2007.
6. Баранов А.Н., Караулов Ю.Н. Русская политическая метафора: материалы к словарю. – М.: ИРЯ, 1991.
7. Башмаков И.А., Башмаков А.И. Интеллектуальные информационные системы. М.: МГТУ им. Н.Э. Баумана, 2005.
8. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.
9. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. СПб.: Книжный Дом, 2007.
10. Березин Ф.М., Головин Б.Н. Общее языкознание. М.: Просвещение, 1979.
11. Болц Н. Азбука медиа / пер. с нем. Л. Ионина, А. Черных. – М.: Европа, 2011.
12. Большакова Е.И. Компьютерная лингвистика: методы, ресурсы, приложения // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. С. 90-105.
13. Бонч-Осмоловская А.А., Толдова С.Ю. Разрешение анафоры // Фонд знаний «Ломоносов». – М., 2011. [www.lomonosov-fund.ru/enc/ru/encyclopedia:0127469:article](http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127469:article) (қаралған күн: 28.02.2012).
14. БЭС — Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998.
15. Воробьева Е.И. Информатизация иноязычного образования: основные направления и перспективы. Архангельск: Поморский университет, 2011.
16. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие для студ., аспирантов, преподавателей филологических факультетов. 2-е изд., испр. – М.: Флинта: Наука, 2007.
17. Гейн А.Г., Сенокосов А.И. Справочник по информатике для школьников. Екатеринбург: У-Фактория, 2003.
18. Егоров А. Слово за слово // Домашний компьютер. 2007. № 9. <http://offline.home.ru/2007/135/334406> (қаралған күн: 28.02.2012).
19. Захаров В.П. Корпусная лингвистика: учеб.-метод. пособие. СПб.: СПбГУ, 2005.

20. Захаров В.П. Информационно-поисковые системы: учеб.-метод. пособие. СПб.: СПбГУ, 2005.
21. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. – М.: Академия, 2004.
22. Клышинский Э.С. Начальные этапы анализа текста // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. – М.: МИЭМ, 2011. – С. 106-140.
23. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – М.: Академия, 2006.
24. Лингвистический энциклопедический словарь / Гл. ред. В. Н. Ярцева. – М.: Сов. энциклопедия, 1990. – 685 с.
25. Марчук Ю.Н. Проблемы машинного перевода. – М.: Наука, 1983.
26. Марчук Ю.Н. Автоматический перевод // Большой энциклопедический словарь. Языкознание. – М.: Большая Российская энциклопедия, 1998. – С. 15.
27. Марчук Ю.Н. Компьютерная лингвистика: учеб. пособие. – М.: АСТ Восток-Запад, 2007.
28. Марчук Ю.Н. Модели перевода. – М.: Академия, 2010.
29. Мельчук И.А. Автоматический синтез // Большая советская энциклопедия. – М.: Советская энциклопедия. 1969-1978. <http://dic.academic.ru/dic.nsf/bse/61319/> Автоматический (қаралған күн: 28.02.2012).
30. Мечковская Н.Б. История языка и история коммуникации: от клинописи до Интернета: курс лекций по общему языкознанию. – М.: Флинта: наука, 2009.
31. Мыркин В.Я. Введение в языкознание. Архангельск: Поморский университет, 2005.
32. Овчинникова И.Г., Углонова И.А. Компьютерное моделирование вербальной коммуникации: учеб.-метод. пособие. – М.: Флинта: наука, 2009.
33. Ожегов С.И. Словарь русского языка. 22-е изд., стер. – М.: Русский язык, 1990.
34. Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28-31 октября 2010 г.) / отв. ред. В.А. Баранов. Уфа; Ижевск: Вагант, 2010. – С. 167-171.
35. Основы теории коммуникации: учебник / под ред. М.А. Василика. – М.: Гардарики, 2007.
36. Потапова Р.К. Новые информационные технологии и лингвистика: учеб. пособие. 2-е изд. – М.: Едиториал УРСС, 2004.
37. Потапова Р.К. Речевое управление роботом: лингвистика и современные автоматизированные системы. – М., 2005.
38. Прикладное языкознание: учебник / отв. ред. А.С. Герд. СПб.: СПбГУ, 1996.
39. Рождественский Ю.В. Лекции по общему языкознанию. – М.: Высшая школа, 1990.

40. Розина И.Н. Педагогическая компьютерно-опосредованная коммуникация: теория и практика. – М.: Логос, 2005.
41. Русский ассоциативный словарь / Ю.Н. Караулов, Ю.А. Сорокин, Е.Ф. Тарасов. Кн. 5: Прямой словарь: от стимула к реакции. – М.: ИРЯ РАН, 1998.
42. Селегей В. Электронные словари и компьютерная лексикография // Ассоциация лексикографов Lingvo. [www.lingvoda.ru/transforum/articles/selegey\\_a1.asp](http://www.lingvoda.ru/transforum/articles/selegey_a1.asp) (қаралған күн: 28.02.2012).
43. Семенов А.Л. Современные информационные технологии и перевод. – М.: Академия, 2008.
44. Сокирко А.В., Толдова С.Ю. Сравнение эффективности двух методик снятия лексической и морфологической неоднозначности для русского языка (скрытая модель Маркова и синтаксический анализатор именных групп). – М., 2005. [www.aot.ru/docs/RusCorporaНММ.htm](http://www.aot.ru/docs/RusCorporaНММ.htm) (қаралған күн: 28.02.2012).
45. Степанов А.Н. Информатика: учеб. пособие. – СПб.: Питер, 2006.
46. Сысоев П.В., Евстигнеев М.Н. Методика обучения иностранному языку с использованием новых информационно-коммуникационных Интернет технологий: учеб.-метод. пособие. – М.: Глосса-Пресс; Ростов н / Д: Феникс, 2009.
47. Сүлейменова Ә.Д., Шәймерденова Н.Ж., Смағұлова Ж.С., Аканова Д.Х. Әлеуметтік лингвистика терминдерінің сөздігі. – Алматы, 2008. – 241 б.
48. Титова С.В. Ресурсы и службы Интернета в преподавании иностранных языков. – М.: Изд-во МГУ, 2003.
49. Толдова С.Ю., Бонч-Осмоловская А.А. Автоматический морфологический анализ // Фонд знаний «Ломоносов». – М., 2011. [www.lomonosovfund.ru/enc/ru/encyclopedia:0127430](http://www.lomonosovfund.ru/enc/ru/encyclopedia:0127430) (қаралған күн: 28.02.2012).
50. Толдова С.Ю., Архипов А.В., Логинова Е.А., Попова Д.П. Корпусная лингвистика // Фонд знаний «Ломоносов». – М., 2011. [www.lomonosovfund.ru/enc/ru/encyclopedia:01210:article](http://www.lomonosovfund.ru/enc/ru/encyclopedia:01210:article) (қаралған күн: 28.02.2012).
51. Тіл білімі сөздігі. Словарь по языкознанию / құраст. Сүлейменова Ә.Д., Мәдиева Г.Б., Шәймерденова Н.Ж., Авакова Р.Ә. – Алматы, 1998.
52. Филиппович Ю., Чернышева М. Историческая компьютерная лексикография – terra incognita в компьютерном мире // Компьютера. 1999. – № 45. <http://offline.computerra.ru/1999/323/3379> (қаралған күн: 13.05.2011).
53. Фролов А.В., Фролов Г.В. Синтез и распознавание речи. Современные решения: электронный учебник. <http://frolov-lib.ru/books/hi/ch00.html> (қаралған күн: 28.02.2012).
54. Хан У., Мани И. Системы автоматического реферирования // Открытые системы. 2000. № 12. [www.osp.ru/os/2000/12/178370](http://www.osp.ru/os/2000/12/178370) (қаралған күн: 02.02.2012).
55. Хомский Н., Миллер Дж. Введение в формальный анализ естественных языков. – М.: Едиториал УРСС, 2003.
56. Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария: практ. руководство для студ., аспирантов, преподавателей филологов. – М.: Флинта: Наука, 2007.

57. Bolshakov I.A., Gelbukh A. Computational Linguistics. Models, Resources, Applications. – Mexico, 2004.
58. Huang C., Simon P., Hsieh S., & Prevot, L. Rethinking Chinese Word Segmentation: Tokenization, Character Classification, or Word break Identification // Proceedings of the Association for Computational Linguistics. Demo and Poster Sessions. Prague, 2007. – P. 69-72. <http://www.aclweb.org/anthology/P/P07/P07-2018.pdf> (қаралған күн: 28.02.2012).
59. Kinnersley B. The Language List. Collected Information On About 2500 Computer Languages, Past and Present. <http://people.ku.edu/~nkinners/LangList/Extras/langlist.htm> (қаралған күн: 28.02.2012).
60. Levy M. CALL: context and conceptualization. Oxford: Oxford University Press, 1997.
61. Maher J. Eliza // Digital Antiquaria, Interactive Fiction. June 15, 2011. <http://www.filfre.net/2011/06/eliza-part-1/> (қаралған күн: 28.02.2012).
62. Matthews C. Intelligent Computer Assisted Language Learning as cognitive science: the choice of syntactic frameworks for language tutoring // Journal of Artificial Intelligence in Education. 1994. № 5/4. – P. 533-556.
63. Rayner M., Carter D. M., Bretan I., Eklund R., Wirén M., Hansen S.L., Kirchmeier-Andersen S., Philp C., Sorensen F., Erdman Thomsen H. Recycling Lingware in a Multilingual MT System // Computation and Language. 1997. [www.aclweb.org/anthology/W/W97/W97-0910.pdf](http://www.aclweb.org/anthology/W/W97/W97-0910.pdf) (қаралған күн: 28.02.2012).
64. Villiger C. Lernsoftware // Angewandte Linguistik: Ein Lehrbuch / Hrsg. Von K. Knapp, G. Antos, M. Becker-Mrotzek u.a. Tübingen; Basel: Francke Verlag, 2004. – P. 187-206.
65. Warschauer M. Computer Assisted Language Learning: an Introduction // Multimedia language teaching / ed. by S. Fotos. Tokyo: Logos International, 1996. – P. 3-20.

## ИНТЕРНЕТ-РЕСУРСЫ

1. Диалог: Международная русскоязычная конференция по компьютерной лингвистике. <http://dialog-21.ru>
2. Лаборатория компьютерной лингвистики Института проблем передачи информации РАН <http://proling.iitp.ru/ru/node/1>
3. Корпусная лингвистика. Машинный перевод. Прикладная лингвистика // Фонд знаний «Ломоносов». <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:01206:article>
4. Корпусная лингвистика: тематический сайт СПбГУ и ИЛИ РАН . СПб., 2008. <http://corpora.iling.spb.ru>
5. Информационные технологии в филологии // Викиверситет. [http://ru.wikiversity.org/wiki/Информационные\\_технологии\\_в\\_филологии](http://ru.wikiversity.org/wiki/Информационные_технологии_в_филологии)
6. Компьютерная лингвистика: научно-образовательный портал «Лингвистика в России: ресурсы для исследователей». [http://uisrussia.msu.ru/linguist/\\_B\\_comput\\_ling.jsp](http://uisrussia.msu.ru/linguist/_B_comput_ling.jsp)

7. Прикладная лингвистика: портал «Единое окно доступа к образовательным ресурсам». [http://window.edu.ru/window/catalog?p\\_rubr=2.2.73.12.15](http://window.edu.ru/window/catalog?p_rubr=2.2.73.12.15)
8. Программы лингвистического анализа и обработки текста. [http://asknet.ru/ Analytics/programms.htm](http://asknet.ru/Analytics/programms.htm)
9. Речевые технологии <http://speech-soft.ru/index.php>
10. Association for Computational Linguistics. <http://www.aclweb.org>
11. Cogprints: free software for Linguistics. University of Southampton. <http://cogprints.org/view/subjects/ling.html>
12. Computational linguistics: MIT Press Journal. <http://www.mitpressjournals.org/loi/coli>
13. Computer-Assisted Language Instruction Consortium. Texas State University. <http://calico.org/>
14. GATES: free software. The University of Sheffield, 1995-2011. <http://gate.ac.uk>
15. Information and Communications Technology for Language Teachers (ICT4LT). Slough, Thames Valley University. [http://www.ict4lt.org/en/en\\_home.htm](http://www.ict4lt.org/en/en_home.htm) 110 111
16. Institut für Computerlinguistik an der Universität Heidelberg. URL: <http://www.cl.uni-heidelberg.de>
17. Language Technology World <http://www.lt-world.org/>
18. LINGUIST List. URL: <http://linguistlist.org>
19. Stanford Engineering Everywhere (SEE): Artificial Intelligence. Stanford University, 1997-2009. <http://see.stanford.edu/see/courses.aspx>

## МАЗМҰНЫ

<b>Кіріспе</b> .....	3
I модуль.	
<b>ТІЛ ЖӘНЕ АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАРДЫҢ НЕГІЗГІ ҰҒЫМДАРЫ</b> .....	5
1-дәріс. Тіл және тіл білімі .....	5
2-дәріс. Ақпарат. Ақпараттық технологиялар .....	13
3-дәріс. Лингвистикадағы ақпараттық технологиялардың аппараттық және бағдарламалық қамтамасыз етілуі .....	25
II модуль.	
<b>АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР: ЛИНГВИСТИКАДА ҚОЛДАНЫЛУ АЯСЫ</b> .....	29
4-дәріс. Дыбыстық сөйлеудің автоматты анализі мен синтезі .....	29
5-дәріс. Мәтінді автоматты түрде тану .....	44
6-дәріс. Мәтінді автоматты түрде аннотациялау және рефераттау .....	46
7-дәріс. Мәтіннің автоматты анализі мен синтезі .....	50
1-аралық бақылау бойынша тест тапсырмалары .....	57
III модуль.	
<b>КОМПЬЮТЕРЛІК ЛИНГВИСТИКА: ҚОЛДАНБАЛЫ САЛАЛАРЫ</b> .....	65
8-9 дәріс. Машиналық аударма .....	65
10-11 дәріс. Компьютерлік лексикография .....	75
12-дәріс. Компьютерлік терминография .....	89
13-дәріс. Тілдерді компьютерлік оқыту .....	91
14-дәріс. Корпустық лингвистика .....	97
15-дәріс. Ақпараттық-іздеу жүйелері .....	102
2-аралық бақылау бойынша тест тапсырмалары .....	106
<b>Әдебиеттер</b> .....	113

Оқу басылымы

Мәдиева Гүлмира Баянжанқызы  
Бектемірова Сәуле Бекмұхамедқызы  
Күзембекова Жанна Жалайырқызы

**ЛИНГВИСТИКАЛЫҚ ЗЕРТТЕУЛЕРДЕГІ  
КОМПЬЮТЕРЛІК ТЕХНОЛОГИЯЛАР**

*Оқу құралы*

Стереотиптік басылым

Редакторы *Г. Рустембекова*  
Компьютерде беттеген және  
мұқабасын безендірген *А. Қалиева*

Мұқабаны безендіруде қолданылған сайт  
<http://www.ostglobal.com/>.

**ИБ № 7580.**

Басуға 31.03.2020 жылы қол қойылды. Пішімі 60x84 <sup>1</sup>/<sub>16</sub>.  
Көлемі 7,25 б.т. Офсетті қағаз. Сандық басылыс. Тапсырыс №2199.

Таралымы 200 дана. Бағасы келісімді.

Әл-Фараби атындағы Қазақ ұлттық университетінің  
«Қазақ университеті» баспа үйі.

050040, Алматы қаласы, әл-Фараби даңғылы, 71.

«Қазақ университеті» баспа үйі баспаханасында басылды.

**«ҚАЗАҚ УНИВЕРСИТЕТІ» баспа үйінің  
жаңа кітаптары**

---

**Жусанбаева С.** Көркем мәтін семантикасы: оқу құралы. – 2014. – 156 бет.  
**ISBN 978-601-04-0885-2**

Оқу құралында көркем мәтін семантикасының қалыптасу мәселелері, көркем мәтіннің мазмұндық-құрылымдық ерекшеліктері, оны түзетін тілдік құралдардың қызметі, көркем мәтінде тілдік бірліктердің мағыналарының құбылулары және т.б. қамтылады.

Оқу құралы жоғары оқу орындары филология мамандықтарында оқитын студенттерге арналған.

**Әлкебаева Д.** Қазақ тілінің прагмастистикасы: оқулық. – 2014. – 247 б.  
**ISBN 978-601-04-0901-9**

Оқулықта қазақ тілі прагмастистикасының зерттеу әдістемесі, ғылыми-теориялық негіздері, прагмастистиканың қалыптасу деректері, когнитивті-ассоциациялық, эстетика-стистикалық, коммуникативті-прагматикалық қызметі, ұлттық тілдік бірліктердің табиғаты және стиль түзуші факторлар мен стилистикалық мүмкіндіктерге қатысты мәселелер қарастырылған.

Ұсынылып отырған оқулық филология, журналистика факультетінде оқитын студенттерге, магистранттар мен PhD докторанттарға, ізденушілерге, жалпы прагмастистика ғылымын зерттеушілерге арналады.

**Тарақов Ә.** Жастар тәрбиесі және дәстүр сабақтастығы: тәрбиелік-әдістемелік құрал. – 2014. – 37 бет.

**ISBN 978-601-04-0665-0**

Аталған кітапшада университетте, факультет пен кафедрада атқарылып жатқан тәрбие үдерісіне қатысты түрлі шараларды және олардың жастарға беретін ықпалын талдап көрсеткен. Автор университетте өткізілген түрлі мәдени шараларға талдау жасайды. Оған өзіндік көзқарасын білдіреді. Кітапшада жастарды ғылымға тарту, өнерге баулу, оларды өнегелі өмір сүру салтына бейімдеу, спортты және басқа да өздеріне пайдалы көсіп түрлерін меңгерту жайында сөз етеді. Тәрбиелік-әдістемелік құралда университеттің әрбір түлегінің ұлтжанды, отансүйгіш, патриот азамат болуы туралы айтылады. Бұл кітапша университет эдвайзерлеріне және магистранттарға, студенттерге арналған.

**Әмір Р., Әмірова Ж.** Қазіргі қазақ тілі, құрмалас сөйлемдер жүйесі: оқулық. – 2014. – 110 б.

**ISBN 978-601-04-0669-8**

Оқулық 2009 жылы жарық көрген «Құрмалас сөйлемдер қазақ тілінің грамматикалық, функционалдық жүйесінде» деген оқу құралы негізінде толықтырылып ұсынылып отыр. Оқулықтың алдына қойған басты мақсаты студенттерге қазақ тілі білімі жөнінде басты теориялық мәлімет берумен қатар, олардың тіліміз жөніндегі компетенциясын мол етіп қалыптастырып, лебіз құру үстіндегі талғамына сай келетін тілдегі ресурстарды ашып тануға, игеруге ұмтылдыру. Әр салаға қатысты теориялық түсініктерді іс-тәжірибеде бекітетін практикалық тапсырмалар, қажетті әдебиеттер тізімі әр тақырыптан кейін берілген. Тапсырмалар теориялық тақырыптарды игеруді, оның практикалық аспектілерін пысықтайтын ұтымды жолдарына байланыстырылып құрылған.

Оқулық студенттер мен магистранттарға, оқытушыларға арналған.

Кітаптарды сатып алу үшін «Қазақ университеті» баспа үйінің сату және маркетинг бөліміне хабарласу керек. Байланыс тел: 8 (727) 377-34-11.

E-mail: [baspa@kaznu.kz](mailto:baspa@kaznu.kz), сайт: [www.read.kz](http://www.read.kz), [www.magkaznu.com](http://www.magkaznu.com)