

МРНТИ16.31.21

А.Ә.Жаңабекова¹, А.А.Қаршығаева²

¹А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің меңгерушісі, филология ғылымдарының докторы.
Алматы қаласы, Қазақстан

²А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің ғылыми қызметкері, филология ғылымдарының кандидаты.
Алматы қаласы, Қазақстан

МӘТІНДЕР КОРПУСЫНА ФОНЕТИКАЛЫҚ БЕЛГІЛЕНІМДЕР ЕНГІЗУГЕ АРНАЛҒАН НҰСҚАУЛЫҚ СИПАТТАМАСЫ (Профессор Ә.Жүнісбектің ұшқадамдық ұстанымы бойынша)

Аннотация: Мәтіндер корпусына енгізілген фонетикалық белгіленім программасы Ә.Жүнісбектің ұшқадамдық жүйесі бойынша сипатталған. Бірінші –сөздің жуан/жіңішкелігіне қарай әуезін (тембр) көрсетеді. Екінші –сөзді автоматты түрде буынға бөледі және қандай буын екенін сипаттайды. Үшінші – сөз құрамындағы әрбір дыбыстың фонетикалық талданымын көрсетеді. Болашақтакорпусқа енгізілген бұл фонетикалық сипаттама буынды автоматты бөле алатын программаның интернет желісінде жұмыс істеуін қамтамасыз етеді.

Мақалада мәтіндер корпусына енгізілген фонетикалық белгіленім бағдарламасын жасауға арналған лингвистикалық нұсқаулыққа сипаттама беріледі. Профессор Ә.Жүнісбектің ұшқадамдық нұсқаулығы басшылыққа алынып, программаға дайындалу барысы қарастырылады. Сонымен қатар лингвистикалық мәлі-меттер енгізілгеннен кейінгі фонетикалық белгіленім терезесінен скриншот беріледі.

Тірек сөздер: мәтіндер корпусы, белгіленім, фонетикалық белгіленім, фонетикалық нұсқаулық, буын, әуез

А.А. Жаңабекова¹, А.А. Каршығаева²

¹заведующий отделом Прикладной лингвистики Института языкознания имени А. Байтұрсынова, док.филол.н., Алматы, Казахстан

²старший научный сотрудник отдела Фонетики Института языкознания имени А. Байтұрсынова, канд.филол.н., Алматы, Казахстан

ОПИСАНИЕ ИНСТРУКЦИЙ ДЛЯ ВНЕДРЕНИЯ ФОНЕТИЧЕСКИХ РАЗМЕТОК В КОРПУС ТЕКСТОВ

Аннотация. Программа фонетической разметки, входящая в корпус текстов, описывается трехступенчатой системой А. Жунусбека. Первый показывает тембр в зависимости от твердости / мягкости слова. Второй – автоматически делит слово на слоги и описывает, какой это слог. Третий показывает фонетический анализ

каждого звука в слове. Это фонетическое описание, которое будет включено в корпус в будущем, позволит программе, которая может автоматически разделять слоги, работать в Интернете. В статье описывается лингвистическая инструкция для разработки программы фонетической разметки, внедренной в корпус текстов. На основе трехшаговой инструкции профессора А.Жунисбека рассматривается процесс подготовки к программе. А также делается скриншот окна фонетической разметки после внедрения лингвистических данных.

Ключевые слова: корпус текстов, разметка, фонетическая разметка, фонетическая инструкция, слог, тембр

A.A. Zhanabekova¹, A.A. Karshygaeva²

¹Head of the Department of Applied Linguistics, A. Baitursynyly Institute of the Linguistics, Doctor of Philology, Almaty, Kazakhstan

²Senior Researcher, Phonetics Department, A. Baitursynyly Institute of the Linguistics, Candidate of Philological Sciences, Almaty, Kazakhstan

DESCRIPTION OF INSTRUCTIONS FOR IMPLEMENTING PHONETIC MARKINGS INTO THE TEXT BODY

Annotation. The phonetic notation program included in the main text is described by A. Zhunisbek's three-stage system. The first shows the melody (timbre) depending on the thickness / subtlety of the word. The second one automatically divides the word into syllables and describes which syllable it is. The third shows the phonetic analysis of each sound in the word. This phonetic description, included in the future case, allows the program to run on the Internet, which can automatically separate syllables.

The article describes the linguistic instruction for the development of a phonetic markup program, embedded in the text corpus. Based on the three-step instructions of Professor A. Zhunisbek, the process of program preparation is considered. And also a screenshot of the phonetic markup window after the introduction of linguistic data is given.

Key words: text corpus, markup, phonetic markup, phonetic instruction, syllable, timbre

Мәтіндер корпусы тілдердегі әртүрлі стильдегі жазба мәтіндердің электронды түрдегі жиынтығы ғана емес, ол тіл деңгейлеріне қатысты тілдік талдауларды автоматты түрде талдап көрсететін компьютерлік бағдарламаға негізделген замануи тілдік құрал. Сондықтан корпустарға осындай автоматты талдаулар жасайтын бағдарламалар жасау керек. Ал компьютер формальді модельдерге сүйеніп қана әртүрлі амалдарды жасай алатындығын ескерсек, онда компьютерлік бағдарламаға лингвистикалық нұсқаулықтар енгізу қажеттігі туындайды. Олай болса, компьютер автоматты түрде тілдік талдауларды жасау мүмкіндігіне ие болу үшін тілдің әр деңгейіне сәйкес лингвистикалық әзірлемелер жасау қолданбалы бағытта лингвистикалық өнімдер дайындайтын тілші-мамандардың алдында тұрған маңызды мәселе.

Қазақ тіл білімінде қазіргі кезде ХХІ ғасырдың бас кезеңдерінен бастап, ұлттық корпустар жасау ісі қолға алынып келеді. А.Байтұрсынұлы атындағы Тіл білімі

институтында Қолданбалы лингвистика бөлімі де ақпараттық технология заманының талабына сәйкес ұлттық корпустар құрастыру ісіне белсенді араласып келеді. Профессор А.Қ.Жұбановтың бастамасымен корпус құрастырудың әдістәсілдері, технологиясы барынша зерттеліп, А.Жұбанов және А.Жаңабекованың авторлығымен «Корпустық лингвистика» оқу құралы ретінде де жарық көрді. Тәжірибе жүзінде 2009 жылдардан бері лингвистикалық белгіленімдер енгізудің бағдарламалары да жасалып, біршама тәжірибе жинақталды. Бұл ретте алдымен морфология деңгейінің талдаулары жүзеге асырылды. Мұндай автоматты бағдарламаны морфологиялық анализатор деп атаймыз. Программалық жағын информатика институтының бұрынғы ғылыми қызметкері Қ.Қойбағаров жасады. Қ.Қойбағаровтың программалық қолдауымен және магаметатик әрі тілші-ғалым, Қолданбалы лингвистика маманы, профессор А.Жұбановтың идеяларымен лингвистикалық және экстралингвистикалық белгіленімдер қою ісі жүзеге асырылуда. Осындай лингвистикалық белгіленімдер әзірлемесін дайындауда тіл білімінің сала мандарының да білік-дағдылары қажет екендігі белгілі. Профессор Ә.Жүнісбек фонетика саласының теориялық мәселелерімен шұғылданып қана қоймай, тіл білімінің қолданбалы бағыттары – оқулық, әдістеме, сол сияқты сөйлеу синтезі және анализі сияқты бағдарламалық жұмыстармен да айналысып отырады. Осы ретте корпуста фонетикалық белгіленімдерді Ә.Жүнісбектің 3-қадамдық нұсқаулығына сүйеніп жасап отырмыз.

Ә.Жүнісбектің 3-қадамдық нұсқаулығында мәтіндегі сөздер үш түрлі белгісі бойынша сипатталады. Олар: сөздің әуезі, сөздерді буынға автоматты бөлу, дыбыстардың (әріптер) фонетикалық сипаттамасы. Осы белгілер бойынша программаға төмендегідей нұсқаулық дайындалды.

Бірінші қадам: Сөздің әуезін (тембр) көрсету. Қазақ тілінде сөздер дауысты дыбыстың жауан/жінішкелігіне қарай жуан әуезді және жінішке әуезді болып бөлінеді. Программа корпуста автоматты түрде сөздердің осы жуан не жінішке екенін ажырып көрсетуі қажет. Қазақ тілінде түбір сөздерге қосымшалар жалғанғанда сөз құрамындағы дауысты дыбыстың жуан/жінішкелігіне қарай жалғанады. Соған сәйкес жуан әуезді не жінішке әуезді сөз болып анықталады. Осы белгіні программа автоматты түрде ажырату үшін мынадай ереже беру керек. Егер сөздің бірінші буынындағы (немесе сөз құрамындағы) дауысты дыбыстар жуан дауыстылар (*а, ы, ұ, о*) болса, онда бұл сөз сингормониялық тұрғыдан *жуан әуезді* болып саналады. Егер сөздің бірінші буынындағы (немесе сөз құрамындағы) дауысты дыбыстар жінішке дауыстылар (*а, ы, ұ, о*) болса, онда бұл сөз сингормониялық тұрғыдан *жінішке әуезді* болып саналады.

Екінші қадам: Сөзді буынға бөлу және оларға сипаттама беру. Қазақ тілінде буын сөздің құрамындағы дауысты дыбыстардың санына қарай анықталады. Сөз құрамында қанша дауысты дыбыс бар, сөзде сонша буын болады. Мысалы:

Әліппе. 3 дауысты дыбыс бар, соған сәйкес 3 буын құрайды: *Ә-ліп-пе*

Ана. 2 дауысты дыбыс бар, соған сәйкес 2 буын құрайды: *А-на*

Бала. 2 дауысты дыбыс бар, соған сәйкес 2 буын құрайды: *Ба-ла*

Жаңбыр. 2 дауысты дыбыс бар, соған сәйкес 2 буын құрайды: *Жаң-быр*

Құттықтады. 4 дауысты дыбыс бар, соған сәйкес 4 буын құрайды: *Құт-тық-та-ды*

Фонетика саласындағы буын теориясы ең күрделі және маңызды мәселелердің бірі деуге болады. Осыған байланысты қазақ тілінің теориялық негізін қалаушы ғалым проф. Құдайберген Жұбанов: «Қазақ тілінің табиғатын дұрыс тану үшін буын

жүйесін жете білу керек. Буынды жақсы білмесең, емлені де білмейсің: сөз тудыру, сөз түрлендіру сияқты морфология негізін де ұға алмайсың, буынды білмейінше, сөз екпіні (ударение), сөз сазы (интонация) секілді сөйлем жүйесіне қатысты тіл құбылыстарының да сырына қана алмайсың, «ажарлап сөйлеу», «эсерлеп оқу» сияқты сабақтарды да дұрыс алып бара алмайсың; буын жайына қанық болмасаң, өлең құрылысының негізімен дұрыс таныс бола алмайсың ...» – дейді [1].

Буын құрылысын сипаттау үшін халықаралық жүйедегі дауысты/дауыссыз дыбыстың белгісін аламыз. Олар: **V(вокализм)** – дауысты дыбыстар; **С (кон-сонантизм)** – дауыссыз дыбыстар. Фонетикалық белгіленім қоюдың компьютерлік бағдарламасы дыбыстарға (әріптерге) қарай буындарды анықтау үшін қазақ әліпбиіндегі әріптерді жоғарыда аталған **V, С** сипатына қарайтізімделіп белгіленуі қажет. Мысалы: **А- V; Ә – V; Ы – V** т.б. Сол сияқты дауыссыз дыбыстар **П – С; Б – С; М – С** т.б.

Қазақ тілінде буынның әдетте 3 түрі бар дейміз: ашық буын, тұйық буын, бітеу буын. А.Байтұрсынұлы сөйлеу мен сөйлем, сөйлем мен сөз, сөз бен буын, буын мен дыбыс желісін бір-бірінен туындатып, арасын үзбей өзара байланыстырып қарайды. Сөздерді буынға бөлудің жолын қарастыра келіп, «Буыншы әріп басқа-лардың ортасында қамауда тұрса, ол буынға «бітеу» деп ат қойындар; буын аяғы буыншы әріпке тірелсе, ол буынға «ашық» деп ат қойындар; буын буыншыдан басқа әріпке тірелсе, ол буынға «тұйық» деп ат қойындар» деп, буын түрлерін саралап береді [2, 344-345 б]. Сондағы «буыншы» әріптеріміз дауысты дыбыстар болып шығады. Буын құраудағы дауысты әріптердің мәнін «әлгі әріптер кіріспейтін буын» болмайды дейді.

Ә.Жүнісбек сөз басында келетін бір дауыстыдан тұратын ашық буынды – дауысты буын деп атайды [3]. Сөйтіп, қазақ тілінде 4 буын түрін көрсетеді: *дауысты, ашық, тұйық, бітеу*.

Буын жігін автоматты түрде табуға қатысты қазақ тіл білімінде алғаш пікір айтқан ғалым Қ.Жұбанов болды. Қ.Жұбановтың «Буын жігін қалай табуға болады?» атты мақаласында буын түрлерінің өрнегін беруде дыбыс аттары мынадай қысқартылған пішінде жазылған: **Ды** – дауысты дыбыс; **Дз** – дауыссыз дыбыс; **сДЗ** – созымды дауыссыз дыбыс; **қДЗ** – қатаң дауыссыз.

Проф. Қ.Жұбанов буынның дыбыс құрамына қарай алты түрін көрсеткен.

- 1) **Ды** (Г) буыны, немесе **жалаң** буын (жалғыз-ақ дауысты дыбыстан жасалады).
- 2) **Дз-Ды** (СГ) буыны, немесе **ашық** буын.
- 3) **Ды-Дз** (ГС) буыны, немесе **жеңіл тұйық** буын.
- 4) **Дз-Ды-Дз** (СГС) буыны, немесе **жеңіл бітеу** буын.
- 5) **Ды-сДЗ-қДЗ** (ГСС) буыны, немесе **ауыр тұйық** буын.
- 6) **Дз-Ды-сДЗ-қДЗ** (СГСС) буыны, немесе **ауыр бітеу** буын.

Осы келтірілген буын түрлерінің алғашқы төртеуі Қ.Жұбановтың пайымдауынша: «... қазақ тілі буындарының негізгі түрлері болып саналады.

Қ.Жұбанов жоғарыдағыдай буындардың моделі арқылы «буын жігін онай, механик түрде табуға болады» – дейді [1, 361 б.].

Қазақ тілші-ғалымдарының буын құрамына, түрлеріне қатысты тұжырымдарын негізге ала отырып, біз корпусқа енгізілетін фонетикалық белгіленім бағдарламасы үшін буын түрлерін төмендегіше сипаттаймыз (буындар қарайтылып берілген):

Дауысты буын - **V** (**а**-на, **о**-тан, **а**-та, **ә**-ліп-пе, **е**-ді, **ө**-мір, **ы**-дыс, **і**-ні т.б.);

Ашық буын - **С+V** (**ба**-ла, **қа**-ла, бөл-**ме**, **ә**-**ке**, **са**-бақ т.б.);

Тұйық буын - **V+С** (**ақ**-ша, **ақ**-та-ды, **ар**-нау, **өр**-ле-ді, **ор**-на-лас-қан);

Тұйық буын - **V+С+С** (**ант**, **айт**-ты);

Бітеу буын - C+V+C (бай-лық, қой-ды, бер-ді, ә-кел-ді, та-быс, ең-бек, жазды т.б.);

Бітеу буын - C+V+C+C (бұлт, қант, жалт, жырт-ты, құрт-ты т.б.);

Бітеу буын- C+C+V +C (Өз-бек-стан, Қа-зақ-стан т.б.).

Программа корпуста сөздерді автоматты түрде бөле алуы үшін осы буындардың тұрақты позициялық шептері туралы нұсқаулық беру керек болады. Мәселен, V дауысты буын (а-на, о-тан, а-та) тек сөз басында ғана кездеседі; C+V ашық буын (са-бақ, бер-ме-ді, бөл-ме)сөздің барлық позицияларында ұшырасады; V+C тұйық буын (ақ-ша, ақ-та-ды, ар-нау) тек сөз басында тұрады; V+C+C тұйық буыны (ант, айт-ты) да тек сөз басында кездеседі; C+V+C бітеу буын(қой-ды, бер-ді, ә-кел-ді, та-быс) сөздің барлық шебінде тұра береді; C+V+C+C бітеу буын(бұлт-ты, жаң-ғырт-ты, жаз-дырт) да барлық жерде тұра береді; C+C+V+C бітеу буыны (Өз-бек-стан, Қа-зақ-стан-да) сөз басында кездеспейді, сөз ортасы мен сөз соңында тұрады.

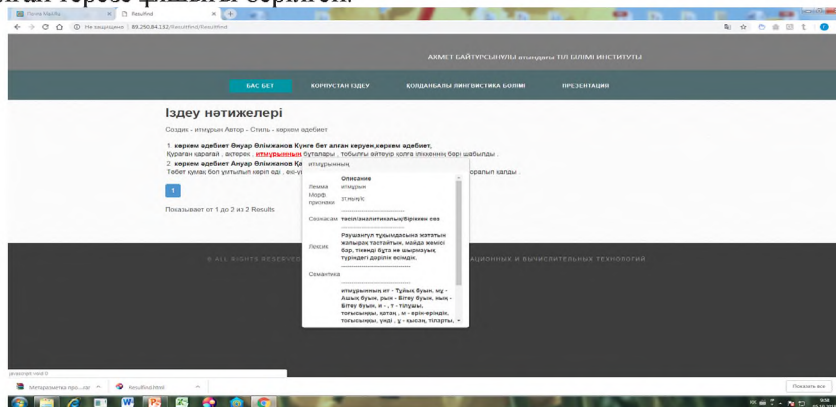
Программа осы нұсқаулық бойынша сөздерді автоматты түрде бөледі. Ал қазақ тіліндегі и мен у-мен келетін кейбір сөздерді бөле алмайды. Мысалы, *баруы, келуі, балуан, алуан* т.б. Өйткені бұл сөздерде у, и дыбстарының алдында тұрған дауыстылар түсіріліп жазылған. Егер бұл сөздер *барууы, келууі, балууан, алууан* болып жазылған болса, автомат мұндайды буынға берілген нұсқаулық бойынша еркін бөле алған болар еді.

Үшінші қадам. Корпусқа фонетикалық белгіленім енгізудегі келесі фонетикалық сипаттама дыбыстардың талданымы. Ол үшін программаға әрбір дыбыстың (әріптің) тілдің, жақтың, еріннің қатысына қарай белгілері сипатталған тізім жасалады. Мысалы: А – ашық, тіларты, езулік; Ә - ашық, тілортасы, езулік; Ы – қысаң, тіларты, езулік; т.б. Біз мұндай сипаттаманы Ә.Жүнісбектің еңбегінен алып отырмыз.

Сонымен, Ә.Жүнісбек корпуста енгізілетін сөз құрамының фонетикалық талданымын мына түрде сипаттайды:

ӘЛШПЕ – жінішке әуезді, **ә-ліп-пе:** **ә** – дауысты буын, **ліп** – бітеу буын, **пе** – ашық буын; **ә** – ашық, тілортасы, езулік; **л** – тілұшы, жанама жуысыңқы, үнді; **і** – қысаң, тілортасы, езулік; **п** – ерін-еріндік, тоғысыңқы, қатаң; **е** – құранды, тілортасы, емеурін езулік.

Төмендегі *1-суретте* корпуста жұмыс істеп тұрған фонетикалық белгіленім қойылған терезе ұяшығы берілген.



1-сурет. А.Байтұрсынұлы атындағы Тіл білімі институты, Қолданбалы лингвистика бөлімі жасаған Мәтіндер корпусындағы лингвистикалық (фонетикалық та) белгіленімдер көрсетілген терезе ұяшығы

Ұяшықтың төменгі жағында сөз фонетикалық тұрғыдан жоғарыда берілген Ә.Жүнісбектің үшқадамдық жүйесі бойынша сипатталған. Бірінші – сөздің жуан/жінішкелігіне қарай әуезін (тембр) көрсеткен. Екінші –сөзді автоматты түрде буынға бөлген және қандай буын екенін сипаттайды. Үшінші – сөз құрамындағы әрбір дыбыстың фонетикалық талданымын көрсетеді. Қорыта келгенде, корпусқа енгізілген бұл фонетикалық сипаттама болашақ буынды автоматты бөле алатын программаның интернет желісінде жұмыс істеуін қамтамасыз етеді. Оның арғы жағында мәтіндік (текстовой) редактордың сөздердің тасымалын тексеруге қатысты қызметтерін жүзеге асыруға,сонымен қатар келешекте лингвистиканың көптеген қолданбалы бағыттағы мәселелерін шешуге септігін тигізбек.

ӘДЕБИЕТТЕР ТІЗІМІ:

- [1] Жұбанов Қ. Қазақ тілі бойынша зерттеулер. – Алматы: Ғылым, 1966.
- [2] Байтұрсынов А. Тіл тағылымы. – Алматы: Ана тілі, 1992.
- [3] Жүнісбек Ә. Қазақ тіл білімінің мәселелері. – Алматы: Абзал-ай, 2018.