

ЎЗБЕК СТЕММЕРИДА СЎЗНИНГ ТОВУШ ЎЗГАРИШИГА УЧРАШИ МАСАЛАСИ

Ботир Болтаевич ЭЛОВ

доцент

техника фанлари бўйича (PhD) фалсафа доктори

Алишер Навоий номидаги ўзбек тили ва адабиёти университети

Тошкент, Ўзбекистон

Зилола Юлдашевна ХУСАИНОВА

таянч докторант

Алишер Навоий номидаги ўзбек тили ва адабиёти университети

Тошкент, Ўзбекистон

Хабиба Абдуманнон қизи БЕРДИЕВА

магистр

Алишер Навоий номидаги ўзбек тили ва адабиёти университети

Тошкент, Ўзбекистон

Аннотация

Ушбу мақолада ўзбек тили корпуси матнларида стеммингни амалга ошириш жараёнидаги сўзларнинг товуш ўзгаришига учраши масаласини ҳал этиш моделлари келтирилган.

Таянч сўзлар: табиий тилни қайта ишлаш, NLP, ўзак, стем, товуш ўзгариши, POS теглаш, морфологик анализатор.

ВОПРОС ЗВУКОВЫХ ИЗМЕНЕНИЙ СЛОВА В УЗБКСКОМ СТЕММЕРЕ

Ботир Болтаевич ЭЛОВ

доцент

доктор философии по (PhD) техническим наукам

Университет узбекского языка и литературы имени Алишера Навои

Ташкент, Узбекистан

Зилола Юлдашевна ХУСАИНОВА

базовый докторант

Университет узбекского языка и литературы имени Алишера Навои

Ташкент, Узбекистан

Хабиба Абдуманнон кизи БЕРДИЕВА

Магистр

Университет узбекского языка и литературы имени Алишера Навои

Ташкент, Узбекистан

Аннотация

В данной статье представлены модели решения проблемы звуковых изменений в процессе стемминга в текстах корпуса узбекского языка.

Ключевые слова: обработка естественного языка, NLP, основа, стемм, изменение звука, разметка, морфологический анализатор.

Ўзбек тилида ўзак ва қўшимча (аффикс)ларни бирлаштириб сўз ҳосил қилинади. Ўзакка аффикслар қўшилганда фонетик уйғунлик ва дисгармония юзага келганлиги сабабли ҳам фонетик, ҳам морфологик ўзгаришларни таҳлил қилиш муҳим вазифа ҳисобланади. Кўпгина NLP вазифаларни ҳал қилишда сўз шакллари уларнинг ўзаккача қисқартириш (стемлаш)га тўғри келади. Сўздан барча флектив аффиксларни олиб ташлаш ва сўзнинг қолган қисмини лемматизация қилиш табиий тилни қайта ишлаш (NLP)нинг муҳим вазифаларидан бири ҳисобланиб, ушбу жараён *стемминг* деб юритилади. Стемминг жараёни ахборот қидириш (IR, Information Retrieval) тизимларида муҳим аҳамият касб этади [6;41-48].

Ўзбек тилида сўзнинг энг кичик маъноли қисми илдиз (*root*) сифатида таърифланган бўлса, *стем* сўзга маъно берувчи энг катта қисмидир. Шунинг учун сўзни икки қисмдан иборат дейишимиз мумкин: *маънони англатувчи стем* ва *қўшимчалар*.

Стем – сўз шаклининг қўшимчаларини кесиб ташлашдан ҳосил бўлувчи қисм бўлиб, баъзи ҳолларда маъно англатмаслиги мумкин. Шунингдек, *стем* сўзнинг морфологик ўзаги билан айнан мос бўлмаслиги ёки мос тушиши мумкин.

Стемминг (ўзбек) – бу сўзга қўшилган *деривацион* ва *флектив* қўшимчаларни олиб ташлаш орқали унинг ўзагигача қисқартириш вазифасидир.

Стемминг жараёнига сўзларнинг қўшимчаларини шунчаки кесиб ташлайдиган “қўпол” эвристик жараён сифатида қаралиши мумкин. Муаллифларнинг фикрига кўра, лемматизациядан фарқли ўлароқ, *стемминг* жараёнида *лугатдан* ёки *морфологик таҳлилдан* фойдаланилмайди [9; 183-188]. Стемминг натижасида ҳосил қилинган ўзак *ҳақиқий сўз* ёки унинг морфологик илдизига ўхшаш бўлиши шарт эмас. Стемминг жараёнининг мақсади – ўхшаш сўзларни бир хил ўзаккача қисқартириш.

Ўзбек тилидаги сўзлар ўзакка баъзи қўшимчалар (аффикслар) қўшиш орқали ҳосил қилинади. Баъзи ҳолларда сўзда фонетик ўзгаришлар юзага

келиши мумкин ва бу бевосита матнда ўз аксини топади. Ўзакнинг ўзи ҳам сўзнинг ўзига хос маъносини ифодаловчи сўз бўлиши мумкин. Аффикслар гапда муҳим рол ўйнаса-да, мустақил маънога эга бўлмайди. Аффикслар сўз ясовчи (derivational suffixes) ва шакл ясовчи (inflectional suffixes) турга ажратилади [10]. Шакл ясовчи қўшимчалар сўзнинг фақат грамматик вазифасини ўзгартиради. Сўз ясовчи қўшимчаларни ўзакка қўшиш орқали сўзда семантик ўзгариш юзага келиши мумкин. Шакл ясовчи қўшимчалар сўзда синтактик ўзгаришларни келтириб чиқаради. Ўзакка аввал сўз ясовчи қўшимчалар, сўнгра шакл ясовчи қўшимчалар бириктирилади [1; 56].

Сўзга қўшила оладиган қўшимчалар сони ва уларнинг кўп сонли бирикмалари тилларда ўзакни аниқлаш жараёнини мураккаб муаммога айлантиради. Чунки кўпчилик агглютинатив тилларда қўшимчалар комбинацияси мураккаб сўз шакллари ҳосил қилади. Янги сўз ясовчи ёки сўз шакллари ҳосил қилувчи қўшимчалар сони бўйича кўрсаткичлар ҳам ҳар хил ҳисобланади.

Ўзбек тилидаги сўзларнинг стемини аниқлаш учун ўзак ва унга бирикадиган барча турдаги қўшимчалар аниқлаб чиқилади. Анъанавий стемнинг алгоритмлари қўшимчалар ва баъзи морфологик қоидаларга асосланган бўлиб, стемнинг жараёни натижасида стемдаги ноаниқлик юзага келиши мумкин. Стемнинг жараёнида одатда сўздаги барча турдаги қўшимчалар олиб ташланади. Лекин бу усулда стемнинг амалга оширилганда, баъзи ҳолларда нотўғри натижа олиниши мумкин [9].

Ўзбек тилидаги сўзларнинг стемини аниқлашда қуйидаги муаммолар юзага келиши мумкин:

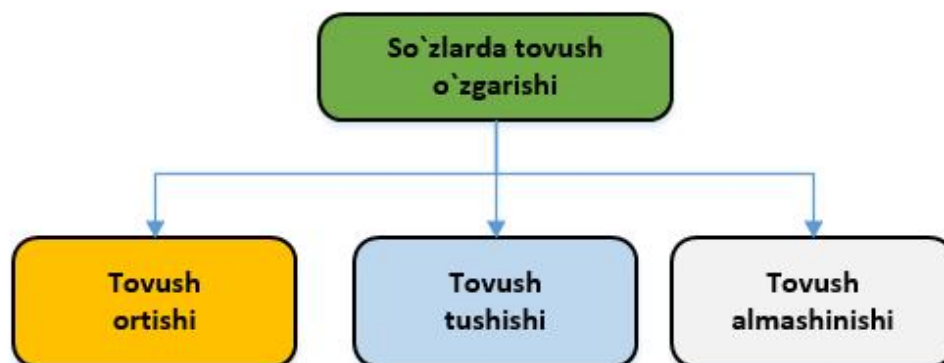
- *ўзак ва қўшимчанинг битта ўзак билан омоним бўлиши;*
- *сўзнинг товуш ўзгаришига учраши;*
- *неологизм ва NERларни стемлаш.*

Ушбу тадқиқот ишида ўзбек тилидаги сўзларнинг стемини аниқлашда сўзнинг товуш ўзгариши билан боғлиқ муаммолар ва уларни ҳал қилиш усуллари келтирилади.

Сўзларда товуш ўзгариши. Фонема нутқда воқеланар экан, унинг варианти нутқий шароит (ёнма-ён келган товушлар ёки қўшимчалар таъсирига учраши) туфайли турлича намоён бўлади. Натижада, товуш сўз моҳиятида бўлмаган хоссаларга эга бўлади. Айрим товуш ёндош товушга мослашади, айримида ўзгариш кучайиб, бошқа товушга алмашиб кетади. Нутқда содир бўладиган ўзгариш товушнинг комбинатор-позицион ўзгариши дейилади [2;65-67]. Бундай ўзгаришларни нутқий фаолият жараёнида жуда кўп учратамиз. Бу ҳолат жарангсиз ундошнинг жаранглига, портловчи ундошнинг сирғалувчи ундошга ёки бир кенг унлининг бошқа бир кенг унлига алмашиши, сўз таркибига кўмакчи морфемаларнинг қўшилиши натижасида айрим товушларнинг тушиб қолиши ёки ортиши тарзида юз беради. Аҳамиятли томони шундаки, юқорида қайд этиб ўтилган ўзгаришларнинг баъзилари ҳам талаффузда, ҳам ёзувда акс этса, баъзилари фақат оғзаки нутқ жараёнидагина юз беради ва ёзувда акс этмайди.

Бўгин, сўз, фраза таркибига кирувчи нутқ товушлари бир-бирига таъсир қилади, натижада товуш ўзгариши юз беради. Нутқ оқимида юз берадиган товуш ўзгариши фонетик процесс дейилади [3; 47-51]. Демак, ҳар қандай товуш ўзгариши нутқ товушларининг бир-бирига таъсири натижасида содир бўлади. Бундай фонетик жараён ва фонетик ҳодисаларнинг батафсил ўрганилиши фонетика ва фонологиянинг мазмун-моҳиятини чуқурроқ ўрганишимизга кенг йўл очади.

Шакл ясовчи қўшимчаларни ўзакнинг охирига қўшиш натижасида баъзи ҳолларда сўзда фонетик ўзгаришлар (insertion, deletion, phonetic harmony, and assimilation) юзага келиши мумкин [1].



1- рasm. Сўзда фонетик ўзгаришлар

Ўзбек тилида сўзда *товуш ортиши*, *тушиши* ва *алмашиши* (weaking, assimilation) каби уч хил фонетик ўзгариш амалга оширилиши мумкин (1-жадвал).

1- жадвал. Ўзбек тилида сўзда *товуш ортиши*, *тушиши* ва *алмашиши*

<i>oishi</i>	<i>tuishi</i>	<i>almashi</i>
obro'y+imiz	me+ning	boshlig'+i
achch+iq	singl+isi	tarog'+ini
shun+day	ko'ngl+im	san+aydi
parvoy+im	bo'yn+i	yurag+im
un+ga	olt+ovlon	qulog'+ing

Стегни аниқлашдаги товуш ўзгаришига учраши муаммосини ҳал қилиш учун, биринчи босқичда ўзак ва қўшимчаларнинг чегаралари аниқланади, иккинчи босқичда эса лемматизация амалга оширилади. Лемматизация натижасида хато ҳосил қилинган стемлар луғатда мавжуд root (ўзак)га ўзгартирилади.

Орфографик қоидалар. Баъзи сўз асосига муайян қўшимчалар қўшилганда асос ёки қўшимча қисмида бир товушнинг бошқа бир товушга алмашиши билан боғлиқ ҳодисалар ҳам кузатилади. Бу ўзгаришлар фақат талаффузда эмас, ёзувда ҳам ўз ифодасини топади.

1. Сўзларда товуш алмашиши.

Товуш алмашиши унлиларга ҳам, ундош товушларга ҳам бирдек тааллуқли ҳисобланади.

1.1. **а** унлиси билан тугаган феълларга *-v*, *-q*, *-qi* қўшимчаси қўшилганда *а* унлиси *о* айтилади ва шундай ёзилади:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
sayla a	-v	Saylov	sayl o
Tara a	-q	Taroq	tar o
sayra a	-qi	Sayroqi	sayr o

1.2. **i** унлиси билан тугаган кўпчилик феълларга -v, -q қўшимчаси қўшилганда бу унли *u* айтилади ва шундай ёзилади:

<i>асос</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
qazi i	-v	qazuv	qaz u
sovi i	-q	sovuq	sov u
o'qi i	-v	o'quv	o'q u

Изоҳ. Лекин **i** унлиси билан тугаган айрим феълларга - q қўшимчаси қўшилганда бу унли *i* айтилади ва шундай ёзилади [<https://github.com/KhZilola/Python-Codes/blob/main/Tovush%20almashishi.docx>]:

<i>асос,</i> <i>стем</i>	<i>қўшимча</i>	<i>сўзшакл</i>
og'ri i	-q	og'ri q
qavi i	-q	qavi q

1.3. **k, q** ундоши билан тугаган кўп бўғинли ва бир бўғинли сўзларга эгалик қўшимчаси қўшилганда **k** ундоши **g** ундошига, **q** ундоши **g'** ундошига айланади ва шундай ёзилади:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
yurak k	{-im,-ing,-si,-i}	yuragim	yurag g
quloq q	{-im,-ing,-si,-i}	qulog'im	qulog' g'
yo'q q	{-im,-ing,-si,-i}	yo'g'im	yo'g' g'

Изоҳ. Лекин кўп бўғинли ўзлашма сўзларга, бир бўғинли кўпчилик сўзларга эгалик қўшимчаси қўшилганда *k, q* товуши аслича айтилади ва ёзилади [<https://github.com/KhZilola/Python-Codes/blob/main/Tovush%20ortishi.docx>]:

<i>асос, стем</i>	<i>қўшимча</i>	<i>сўзшакл</i>
zav q	{-im,-ing,-si,-i}	zav qi
par k	{-im,-ing,-si,-i}	par ki
no k	{-im,-ing,-si,-i}	no ki

1.4. Баъзи сўзларга *-a, -ay, -la* каби сўз ясовчи қўшимчалар қўшилиши натижасида товуш алмашиши кузатилади:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
ot	-a	ata	at
ong	-la	angla	ang
sariq	-ay	sargʻay	sargʻ

2. *Сўзларда товуш ортиши.* Товуш ортиши деб юритилувчи навбатдаги ходисанинг фақат талаффузда акс этадиган ва ҳам талаффузда, ҳам ёзувда ифода этиладиган икки хил кўриниши мавжуд. Оғзаки нутқ жараёнидагина юзага чиқадиган ҳодиса, асосан, ўзлашма сўзлар доирасида юз беради: *рус→ўрис, икаф→ишкоф, стакан→истакан, трактор→тирактор* каби [7; 12].

Ҳозирги ўзбек адабий тили дарслигида юқоридаги мисолларда кузатилган ўзгаришлар протеза ҳодисасига мисол бўлиши айтилган [1; 244]. Бу ҳодиса туркий, шу жумладан, ўзбек тилида сўз бошида ундош товушларнинг ёнма-ён келиши жуда кам учраши, шу боис ҳам талаффуз қулайлигига эришиш мақсадида улардан аввал бир унлининг орттириб талаффуз қилиниши билан изоҳланади. Ўзбек тилшунослигида товуш ортишининг бир нечта кўринишлари мавжуд бўлиб, улар тилшунос олимлар томонидан атрофлича тадқиқ этилган.

Ўзбек тилида сўзда товуш ортиши қоидалари

Товуш ортиши тури	Қандай содир бўлади?	Мисоллар
Протеза	Сўз бошида битта унлининг қўшиб талаффуз қилиниши билан боғлиқ ҳодиса бўлиб, одатда, сонор [r] товушидан олдин [o'] унлиси орттирилади.	$ro'mol \rightarrow [o'ramol]$, $ro'za \rightarrow [o'raza]$, $rais \rightarrow [o'rais]$, $rang \rightarrow [o'rang]$
Эпентеза	Сўз боши, ўртаси ва охирида икки ундош қатор келганда, улар орасида [i], баъзан [u] ва [a] унлиси орттирилиб талаффуз қилинади.	$fikr \rightarrow [fikir]$, $hukm \rightarrow [hukum]$, $doklad \rightarrow [dakalad]$, $klass \rightarrow [kilass]$

2.1. *у, бу, шу, ўша* олмошларига *-да, -дан, -дай, -даги, -га, -гача, -ча* қўшимчалари қўшилганда *n* товуши қўшиб айтилади ва шундай ёзилади:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
u	-ga	u ng a	un
bu	-ga	b u ng	bun
shu	-dagi	sh u ndagi	shun
shu	-ga	sh u ng	shun

2.2. *parvo, obro', mavqe, mavzu, avzo* сўзларига I, II шахс эгалик қўшимчалари қўшилганда бир *у* товуши қўшиб айтилади ва шундай ёзилади:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
parvo	-im	Parvo y im	parvoy
obro'	-im	obro' y im	obro'y
mavqe	-im	mavqe y im	mavqey
mavzu	-im	mavzu y im	mavzuy
avzo	-im	avzo y im	avzoy

Изоҳ: *Xudo, mavzu* сўзларига эса III шахс эгалик шакли *-si* шаклида қўшилади, ҳеч қандай товуш ортиши кузатилмайди: *xudos*i*, mavzusi*

Протеза ҳақида бошқа адабиётларда ҳам бир қанча маълумотларни учратишимиз мумкин. Сўзнинг бошланғич қисмига товушларни қўшиб айтиш протеза дейилади. Туркий тилларда, жумладан, ўзбек тилида рус тилидан кирган сўзлар бошида келган ундош товушлар бирикмасидан олдин [i], [u] унлилари қўшиб талаффуз қилинади: *stansiya*→[i*stansiya*], *stol*→[u*stol*] каби. Лекин бу ҳодиса камайиб бормоқда [3; 50].

Айрим ҳолатларда сўз бошида сирғалувчи ва портловчи икки ундош қатор келганда [i] унлиси ҳам орттирилиши мумкин: *shkaf*→[i*shkaf*], *spravka*→[i*spravka*], *stol*→[i*stol*], *stul*→[i*stul*], *shtraf*→[i*shtraf*] каби.

3. Сўзларда товуш тушуши. Нутқ товушлари нутқни таркиб топтирувчи материалдир. Нутқ ўз қурилишида мана шу материални тежаб сарфлайди. Нутқ товушларининг нутқ кечимидаги мана шу тежами унинг тушиши, яъни, диереза ҳодисаси деб қаралади [2; 66].

3.1. *O'rin, qorin, burun, o'g'il, bo'yin, ko'ngil* каби баъзи сўзларга эгалик қўшимчаси қўшилганда иккинчи бўғиндаги унли айтилмайди ва ёзилмайди:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
<i>o'rin</i>	-i	<i>o'rni</i> (<i>o'rini</i>)	<i>o'rn</i>
<i>qorin</i>	-i	<i>qorni</i> (<i>qorini</i>)	<i>Qorn</i>
<i>burun</i>	-i	<i>burni</i> (<i>buruni</i>)	<i>Burn</i>
<i>o'g'il</i>	-i	<i>o'gli</i> (<i>o'g'ili</i>)	<i>o'g'l</i>
<i>bo'yin</i>	-i	<i>bo'yni</i> (<i>bo'yini</i>)	<i>bo'yn</i>
<i>ko'ngil</i>	-i	<i>ko'ngli</i> (<i>ko'ngili</i>)	<i>ko'ngl</i>

3.2. *Qayir, ayir* каби феълларга нисбат шаклини ясовчи **-il** қўшимчаси қўшилганда иккинчи бўғиндаги унли айтилмайди ва ёзилмайди:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
<i>qayir</i>	-il	<i>qayril</i> (<i>qayiril</i>)	<i>qayr</i>
<i>ayir</i>	-il	<i>ayril</i> (<i>ayiril</i>)	<i>ayr</i>

3.3. Ikki, olti, yetti сўзларига *-ov*, *-ala*, *-ovlon* қўшимчалари қўшилганда иккинчи бўғиндаги унли айтилмайди ва ёзилмайди:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
ikki	-ov	ikk ov (ikk io v)	ikk
olti	-ala	olt ala (olt iala)	olt
yetti	-ovlon	yett ovlon (yett io vlon)	yett

3.4. Баъзи сўзларга *-a*, *-ay* сўз ясовчи қўшимчаси қўшилганда иккинчи бўғиндаги унли айтилмайди ва ёзилмайди:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
o'yin	-a	o'y na (o'y ina)	o'yn
ulg'	-ay	ulg' ay (ulg' ay)	ulg'
sariq	-ay	sarg' ay (sar iq ay)	sarg'

Изоҳ. Sarg'ay сўзида sariq + ay = sarg'ay *q – g* га товуш алмашиши ҳодисаси ҳам кузатилган.

3.5. Men, sen олмошларига *-ni*, *-ning*, *-niki* қўшимчалари қўшилганда қўшимчадаги n товуши айтилмайди ва ёзилмайди:

<i>асос (root)</i>	<i>қўшимча</i>	<i>сўзшакл</i>	<i>стем</i>
men	-ni	men i (men ni)	men
sen	-ning	sen ing (sen ning)	sen
men	-niki	men iki (men niki)	men

Турли ҳолатларда товушларнинг талаффузида ҳар хил ўзгаришлар юз беради. Н.С.Трубецкойнинг фикрича, амалий жиҳатдан айнан бир товушни бир позицияда ҳам бир неча марта аниқ ва бир турда талаффуз этиш мумкин

эмас [9; 48-49]. Яъни, нутқ эгаси ўз сўзларини турли вазиятларда ҳам, бир хил вазиятларда ҳам хилма-хил оҳангларда талаффуз этади. Шу боис ҳам ҳар сафар нутқ товушлари бошқа-бошқа оҳангда, бошқа-бошқа баландлик ва сифат кўрсаткичлари билан талаффуз этилади. Битта товуш ҳеч қачон айнан ўзидек такрорланмайди. Нутқда товушларнинг позицион ўзгаришлари қанча чуқур ўрганилса, товуш турлари шунча кенг аниқланиши ва бу борада кенгроқ тасаввурга эга бўлиш мумкин.

Morfologik analizator

Matnni kiriting

saylov, taroq, sayroqi.
yuragim, qulog'im.
zavqi, parki, noki.
o'rni, burni, bo'yni.
meni, sening, meniki

Analiz

№	So'z shakli		Lemma		Stem		O'zak		Asos va qo'shimchalar
	Qiymati	So'z turkumi	Qiymati	So'z turkumi	Qiymati	So'z turkumi	Qiymati	So'z turkumi	
1	saylov		saylov	Ot ...	saylo	-	sayla	Fe'l	{saylov}
2	taroq		taroq	Ot ...	taro	-	tara	Fe'l	{taroq}
3	sayroqi		sayroqi	Sifat ...	sayro	-	sayra	Fe'l	{sayroqi}
4	yuragim	Ot	yurak	Ot ...	yurag	-	yurak	Ot	{yurak}-im
5	qulog'im	Ot	quloq	Ot ...	qulog'	-	quloq	Ot	{quloq}-im
6	zavqi	Ot	zavq	Ot ...	zavq	Ot	zavq	Ot	{zavq}-i
7	parki		park	Ot ...	park	Ot	park	Ot	{park}-i

Matnni kiriting

saylov, taroq, sayroqi.
yuragim, qulog'im.
zavqi, parki, noki.
o'rni, burni, bo'yni.
meni, sening, meniki

Stemming

saylo ([saylo]-v) taro ([taro]-q) sayro ([sayro]-q-i) yurag ([yurag]-im) qulog' ([qulog']-im) zavq ([zavq]-i) park ([park]-i) ook ([nok]-i) o'rn ([o'rn]-i) burn ([burn]-i) bo'yn ([bo'yn]-i) men ([men]-i) sen ([sen]-ing) men ([men]-ik-i)

2-расм. Ўзбек тили морфологик анализатори (стемминг).

2-расмда ўзбек тили морфологик анализатори дастурий таъминотининг “*Stemming*” қисмида юқорида келтирилган усуллар жорий этилган бўлиб, товуш ўзгариши ҳодисаси билан боғлиқ баъзи намуналар келтирилган.

Шундай қилиб, юқорида ўзбек тили корпуси матнларида стеммингни амалга ошириш жараёнида сўздаги товуш ўзгариш масаласи кўриб чиқилди, ўзбек тилидаги баъзи сўзлардаги *товуш ортиши, тушиши* ва *алмашиниши* каби уч хил турдаги фонетик ўзгаришлар грамматик қоидалар асосида таҳлил

қилиниб, намуналар воситасида ёритилди. Стемни аниқлашдаги товуш ўзгаришига учраши муаммосини ҳал қилиш учун, биринчи босқичда ўзак ва кўшимчаларнинг чегаралари аниқланиб, иккинчи босқичда эса лемматизация амалга оширилади. Лемматизация натижасида хато ҳосил қилинган стемлар луғатда мавжуд root (ўзак)га ўзгартирилади. Мақолада келтирилган усуллар Б.Элов, Р.Алаев, Ш.Хамроева ва З.Хусаинова томонидан ишлаб чиқилган ўзбек тили морфологик анализаторига татбиқ этилган.

Фойдаланилган адабиётлар:

1. Миртожиев М. Ўзбек тили фонетикаси. – Тошкент: Фан, 2013. – 307 б.
2. Сайфуллаева Р., Менглиев Б. ва бошқ. Ҳозирги ўзбек адабий тили. – Тошкент: 2009. – 298 б.
3. Содиқов А. ва бошқ. Тилшуносликка кириш. –Тошкент: Ўқитувчи, 1981. – 203 б.
4. Трубетской Н.С. Основы фонологии. – Москва, 1960. – 345 с.
5. Элов Б.Б., Ҳамроева Ш.М., Хусаинова З.Ю. NLP (табиий тилга ишлов бериш)нинг pipeline конвеери //Муҳаммад ал-Хоразмий авлодлари. Илмий – амалий ва ахборот – таҳлилий журнали, 2023. – №1(23). – Б.181-192.
6. Элов Б.Б., Ҳамроева Ш.М., Абдуллаева О.Х., Хусаинова З.Ю., Худайбергандов Н.У. Ўзбек, турк ва уйғур тилларида POS теглаш ва стемминг //Ўзбекистон: тил ва маданият (Компьютер лингвистикаси), 2023. – №1(6). – Б. 41-60.
7. Ўзбек тили практикуми. 1-қисм. – Тошкент: 2005. – 278 б.
8. Ҳожиев А. Ўзбек тилида сўз ясалиши. – Ташкент, 2005. – 271 б.
9. Paice, C. D. (1990). Another Stemmer. *ACM SIGIR Forum*, 24(3). <https://doi.org/10.1145/101306.101310>
10. Sharma, A., Kumar, R., & Mansotra, V. (2016). Proposed Stemming Algorithm for Hindi Information Retrieval. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, 3297(6). <https://doi.org/10.15680/IJIRCCE.2016>

11. <http://uznatcorpara.uz/uz/Stemmer>