

НӨХЦӨЛИЙН ЗАЛГАГДАХ ШИНЖЭЭР ҮГИЙГ БҮТЦЭЭР ЗАДЛАХ

Д.Оргил^{1,2*}, Ц.Өнөрбаян³

1-Инженер Технологийн Сургууль, ХААИС,

2-Компьютерийн Ухааны Сургууль, ӨМИС,

3-Монгол Судлалын Төв, МУБИС,

*Email: orgilid@yahoo.com

ХУРААНГУЙ

Монгол хэлний өгүүлбэрийн бүтцийн судалгааг хийхийн тулд үгийн дотор илэрч байгаа хэлзүйн утгыг тодорхойлох шаардлагатай болдог. Мөн хэлзүйн хувиллаар хувилсан үг бүрийг үгийн санд хадгалах нь залгамал бүтэцтэй монгол хэлэнд тохирохгүй байв. Ийм учир үгийг бүтцээр задлахдаа үгийн сангийн болон хэлзүйн утгыг нь ялгах зорилго тавилаа. Гэтэл орчин цагийн кирилл бичигт зөв бичгийн дүрмийн алдаатай бичвэр их тохиолдож байв. Иймээс зөв бичгийн дүрмийн алдаатай бичвэрээс үгийн сангийн болон хэлзүйн утгыг нь ялгах шаардлага тулгарсан. Ийм учраас нөхцөлийн залгаагдах шинжээр үгийг бүтцээр задлах аргыг шинээр зохиолоо. Энэ аргаар зөв бичгийн дүрмийн алдаатай бичигдсэн үгээс үгийн сангийн болон хэлзүйн утгыг ялгаж чадаж байна. Мөн ижил хэлбэртэй үгсийг нөхцөлийн залгаагдах шинжээр нь нэр, үйл эсэхийг ялаж байв. Гэтэл нэр үг дотроо ялгаатай хувилж үгийн утга ялгадаг. Үүнийг ялгахгүй байсан. Иймээс морфотактикийг үгийн салаа утгыг ялгах, зөв бичгийн дүрмийн алдааг засварлахад ашиглав. Хувилсан үгээс үгийн салаа утга ялгах нь чухал ач холбогдолтой. Учир нь кирилл бичигт үг хувилсан хэлбэрээрээ тохиолдох нь элбэг (68%) байдаг. Мөн зөв бичгийн дүрмийн алдааг дүрэм ашиглан автоматаар засварлах ажил хараахан хийгдээгүй шинэ ажил юм. Кирилл бичгийн өнөөгийн цахим боловсруулалтаас харахад монгол хэлний өөрийнх нь онцлогт суурилсан үг хувилах загвар боловсруулах шаардлагатай байв. Иймээс бид кирилл бичгийн хэлзүйн дүрэм, авиазүй зарчим, монгол хүний сэтгэхүйн онцлог зэргийг үг хувилах загвартаа тусгахыг хичээлээ.

ТҮЛХҮҮР ҮГ: Морфотактик, зөв бичгийн дүрмийн алдаа, үгийн салаа утга

ОРШИЛ

Хүний хэлийг бүтцийн хувьд хоёр ангилдаг. Хэлзүйн утга нь гол төлөв үгийн дотор илэрдэг хэлийг нийлэг бүтэцтэй хэл гэж нэрлэдэг. Ийм хэв шинжийн хэлэнд үгийн сангийн болон хэлзүйн утгууд нэг үгэнд нэгдэн цогцолсон байдаг. Гэтэл нөгөө талаас эдгээр хэлзүйн утга нь өөр хэлэнд үгийн гадна, тухайлбал, өгүүлбэрт буюу холбоо үгэнд илэрч болдог бөгөөд тийм хэлийг задлаг бүтэцтэй хэл гэдэг. Иймэрхүү хэв

шинжийн хэлэнд үг нь гол төлөв зөвхөн үгийн сангийн утгыг дамжуулагч болж, харин хэлзүйн утга нь, тусгай биеэ дааж өгүүлбэр дэх үгсийн дэс дараа, холбох ба туслах үгс, аялга гэх мэт хэлзүйн хэрэглүүрээр илэрдэг. Аливаа хэлэнд нийлэг ба задлаг бүтцийн аль нэг нь голлох бөгөөд аль нэг нь огт байхгүй хэл гэж байдаггүй [1]. Монгол хэлэнд нийлэг хэрэглүүр давамгайл хөгжсөн байдаг. Ийм учраас өгүүлбэрийн бүтцийн

судалгааг хийхийн тулд үгийн дотор илэрч байгаа хэлзүйн утгыг тодорхойлох шаардлагатай байдаг. Мөн хэлзүйн хувиллаар хувилсан үг бүрийг үгийн санд хадгалах нь монгол хэлэнд тохирохгүй. Учир нь жинхэнэ үгээсээ тухайн үгийн хэлзүйжсэн хэлбэрүүд нь хэд дахин олон болж үгийн сангийн хэмжээ хэт нүсэр болдог. Иймээс хэлзүйн утгат хэсгийг нь хасаж, үгийн сангийн утгат хэсгийг нь ялган авах шаардлагатай байдаг. Ийм учир үгийг бүтцээр задалж, үгийн сангийн болон хэлзүйн утгыг нь ялгаж тодорхойлох зорилго тавилаа. Кирил үсгийн дүрмийг анх 1942 онд хэвлэж, 1946 онд дахин засварлаж гаргасан бөгөөд тухайн үед зарим дүрмийг тодруулах зорилго агуулж байсан нь дүрэм бүхэлдээ төгс зөв болоогүйн гэрч юм [2]. Уг дүрэмд монгол хэлний авианы байрлалын

хуулийг бүрэн тусгаагүйн улмаас олон зохиомол дүрэмтэй, ялангуяа үгийн язгуур, үндэс, дагавар, нөхцөлийн бүтцийг эвдсэн тохиолдол олон байна [3]. Үүнээс үүдэлтэй зөв бичгийн дүрмийн алдаатай бичвэр их тохиолдох болжээ. Иймээс зөв бичгийн дүрмийн алдаатай бичвэрээс үгийн сангийн зөв утгыг нь олох шаардлага тулгардаг. Үгийг бүтцээр задлахдаа зөв бичгийн дүрмийн алдаатай бичвэрээс үгийн сангийн зөв утгыг нь олох, үгийг хувилгахдаа зөв бичгийн дүрмийн алдааг автоматаар засварлах боломжтой загвар боловсруулахыг оролдов. Зөв бичгийн дүрмийн алдаа шалгах програмыг Болорсофт ХХК статистик арга ашиглан боловсруулсан байдаг. Харин зөв бичгийн дүрмийн алдааг дүрэм ашиглан автоматаар засварлах ажил хараахан хийгдээгүй.

СУДАЛГААНЫ АРГА ЗҮЙ

Үгийг хувилгахдаа үгийн үндсэнд нөхцөл залгах, нөхцөлийн дараа нөхцөл залгах гэсэн хоёр дүрэм болгож хуваав. Учир нь нөхцөлийг үгийн үндсэнд залгахад дан ганц алгоритмаар шийдэж болдоггүй. Зарим үг нийтлэг бус хувилах тохиолдол байх тул үгийн санд морфотактик (үндэс + нөхцөл) үүсгэх шаардлагатай болдог. Харин үгийн бүтцэд оролцож байгаа нөхцөлүүд нарийн тогтсон дараалалтай байдаг тул

нөхцөлийн дараа нөхцөл залгах дүрмийг алгоритмаар шийдэв. **Үгийн үндсийн өөрчлөлт** Нөхцөлийг үгийн үндсэнд залгахад кирил үсгийн зөв бичих дүрмийн дагуу үгийн үндсийн балархай эгшиг гээгдэх, зөөлний тэмдэг “и” болж хувирах зэргээр үндсийн бүтэц өөрчлөгддөг. Хүснэгт 1-д үгийн үндсийн нийтлэг өөрчлөлтүүдийг харуулав.

Хүснэгт 1

Үгийн үндсийн нийтлэг өөрчлөлт

	дарга	аймаг	сахь
1	дарг+ыг	аймг+ийн	сахи+ж

Үүнд: 0-ээр тухайн үг үндсээрээ байх үеийн хэлбэрийг, 1-ээр үгийн үндсийн балархай эгшигийг гэсэн, зөөлний тэмдэг “и” болсон хэлбэрийг тэмдэглэв. Зарим үгийн үндсэнд нөхцөл залгахад кирил үсгийн зөв бичих дүрмээс үл хамаарч үгийн үндэс өөрчлөгддөг. Жишээлбэл

“би” гэдэг үг “мин, над, намай” гэсэн олон янзын хувилбартай болсон нь монгол хэлний түүхэн хөгжлийн явцад гарч ирсэн авиа зүйн үзэгдэл юм. Үгийн үндэс дараах байдлаар нийтлэг бус өөрчлөгддөг

Хүснэгт 2

Үгийн үндсийн нийтлэг бус өөрчлөлт

	би	чи	тэр
1	мин+ий	чин+ий	түүн+ий
2	над+ад	чам+д	түүг+ээр
3	намай+г	чамай+г	

Хүснэгт 1, 2-оос харахад үгийн үндэс 0 хэлбэрээрээ биеэ даасан үг болно. Харин 0-ээс бусад хэлбэрээрээ биеэ даасан үг болохгүй

бөгөөд зөвхөн нөхцөлөөр хувилсны дараа сая биеэ даасан үг болж чадна.

Морфотактик (үндэс + нөхцөл) Ц.Дамдинсүрэн, Б.Осор нар монгол хэлний толь хийхдээ "нэр, үйл үгсийг хэдэн хэлбэрт хувааж, хэлбэр бүрийг тоогоор дугаарлан, тэр дугаар нь тухайн үг яаж нөхцөл авч хувирахыг хавсралт жагсаалтаас олж үзэж болно. Ингэж нэр ба үйл үгийн олон хэлбэрийг бүрэн жагсаахгүй, тэр олон хувилбарыг эцсийн жагсаалтаас үзэхээр хийсэн нь уг толийг авсаархан болоход тус боллоо" гэжээ [4]. Иймээс үгийн үндсийг кирил үсгийн зөв бичих дүрмийн дагуу ижил хувилах эсэхээр

нь бүлэг болгон хуваах санаа бол уг толийг ашиглахад авсаархан болгох үүднээс юм. Харин цахим боловсруулалтад үгийн үндсийг ингэж бүлэг болгон хуваах шаардлагагүй. Иймээс бид кирил үсгийн зөв бичих дүрмийн дагуу үг хэрхэн хувилахыг судалж үзээд үгийн төгсгөлийн үсгээс хамаарсан үг хувилах дүрэм боловсрууллаа. Ингэж үг хувилах дүрэм боловсруулахдаа үгийн бүтцэд тохиолдож болох нөхцөлийн хэлбэр бүрийг бүртгэв (Хүснэгт 3).

Хүснэгт 3

Нэр үгийн хувилал												
Үгийн төгсгөлийн үсэг												
ХЦӨЛ	Гийгүүлэгч		“Ь”		Эгшиг					Зарим нөхцөлийг залгах		
	9	7	9ь	7ь	“и”	“я, е, ё”	“а, э, о, ө”	УЭ	“й”	Уг”н”	“н”	“Г, К, Ж, Ч, Ш”
Fc 1	0 1+ын ийн		1+йн		0+йн	0+ын ийн	1+ын ийн	0+ны ний	0+	0+гий	0 1+ы и	0 1+ийн
Fc 3	0 1+ыг ийг		1+йг		0+йг	0+ыг ийг	1+ыг ийг	0+г		0+г		0 1+ийг
Fr 2	0 1+22д		1+2д		0+2д		1+22д	0+н22д		0+г22д		
Fc 4	0 1+44с		1+4с		0+4с			0+н44с		0+г44с		
Fc 5	0 1+44р		1+4р		0+4р			0+г44р		0+г44р		
Fe 1	0 1+44		1+4		0+4						“Г, 5с”	“В, р”
Fc 2	0 1+5д	0+д	1+д	0+д	0+д			0+нд	0+д		0+т	0+д т
Fc 6			1+т4й									
Fc 1	0+ны ний		1+ны		0+ны	0+ны ний						
Fr 2	0+н22д		1+н22д		0+н22д							
Fc 4	0+н44с		1+н44с		0+н44с							
Fc 2	0 1+5нд		1+нд		0+нд							

Үүнд: 0, 1-ээр үгийн үндсийг, 9-өөр заримдаг гийгүүлэгчийг, 7-оор эгшигт гийгүүлэгчийг тэмдэглэв. Үгийн үндсэнд нөхцөлийн хэлбэрийг ялгаж залгахад "ий, ы" эгшиг хэрхэн өөрчлөгдөхийг тусгах шаардлагатай байв. Учир нь "ий, ы" эгшигийг үгийн үндсэнд ялгаж залгахад үгийн утга өөрчлөгдөж болно. Жишээлбэл "багийн, багын" хоёр үг ялгаатай утгатай. Харин "а, э, о, ө" ба "у, ү" эгшиг үсэг хэрхэн

өөрчлөгдөхийг тусгах шаардлагагүй. Ийм учраас нөхцөлийн санд нөхцөлийн хэлбэрийг үүсгэхдээ "а, э, о, ө" үсгүүдийг 4-ийн тоогоор, "у, ү" үсгүүдийг 2-ын тоогоор, балархай эгшигийг 5-ын тоогоор тус тус бүлэглэж тэмдэглэв. Хүснэгт 3-д харуулсан загварын дагуу морфотактик (үндэс + нөхцөл) хэрхэн үүссэнийг зарим нэр үгийг жишээн дээр харуулав.

Хүснэгт 4

Морфотактик

Үгийн үндэс	Үгийн аймаг	Морфотактик (Үндэс + нөхцөл)
<i>бодол</i>	Nc0	1+Fp21,1+Fc111,0+Fc211,1+Fc311,1+Fc41,1+Fc51,0+Fc61,1+Fe11
<i>бургас</i>	Nc1	0+Fp24,0+Fc161,0+Fc222,0+Fc311,0+Fc44,0+Fc51,0+Fc61,0+Fe11
<i>дарга</i>	Nc0	1+Fp21,1+Fc111,0+Fc211,1+Fc311,0+Fc42,0+Fc52,0+Fc61,0+Fe12
<i>чарга</i>	Nc1	0+Fp24,0+Fc161,0+Fc221,1+Fc311,0+Fc44,0+Fc52,0+Fc61,0+Fe12
<i>үе</i>	Nt9	0+Fp22,0+Fc112,0+Fc211,0+Fc312,0+Fc42,0+Fc52,0+Fc61,0+Fe12
<i>үе</i>	Nc1	0+Fp24,0+Fc162,0+Fc221,0+Fc312,0+Fc44,0+Fc52,0+Fc61,0+Fe12

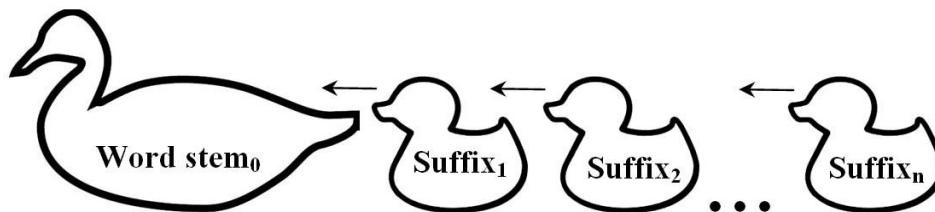
Үүнд: Нөхцөлийн хэлбэр бүрийн хэлзүйн утгыг тэмдэглэхдээ GB/T 26235 - 2010 стандартыг баримталлаа [5].

Үгийг бүтцээр задлах

Монгол хэлний нөхцөлүүд үгийн бүтцэд дараалан орохдоо өмнөх өмнөхөө дагасан бүтэцтэй байдаг.

Өөрөөр хэлбэл нөхцөлийг залгаж болох эсэх нь өмнөх гишүүнээсээ хамааралтай байдаг. Энэ бол залгамал бүтцийн онцлог юм. Иймээс монгол хэлний морфотактикийг галуун цуваатай зүйрлэв.

(1-р зураг).



1-р зураг. Монгол хэлний үгийг галуун цуваатай зүйрлэх нь

Иймээс морфотактикийг судлахдаа ямар нөхцөлийг ямар ямар нөхцөлийн араас залгаж болох вэ? гэдэг талаас нь судлав.

1-р зурагт харуулсны дагуу нөхцөлийн хэлбэр бүрт залгагдах шинж үүсгээд ямар гишүүнтэй

залгагдаж болохыг нөхцөлийнх нь шинжид зааж өглөө. Жишээлбэл, гарах тийн ялгалын нөхцөлийн хэлбэрүүдийг нөхцөлийн санд хэрхэн үүсгэснийг харуулав (Хүснэгт 4).

Хүснэгт 5

Нөхцөлийн сан

F	Нөхцөлийн хэлбэр	Үгийн төгсгөлийн үсэг	Нөхцөлийн залгагдах боломж
Fc4	Гарах тийн ялгал		
Fc41	+44с	Гийгүүлэгч	N, Fp, Fh, Fd1, Fd2, Fd4
Fc42	+4с	"а, э, о, ө, и, я, е, ё"	N
Fc43	+г44с	"н", урт эгшиг, "й"	N
Fc44	+н44с		N

Үүнд: Нөхцөлийн бүх хэлбэр нэр үг (N)-тэй залгагдаж болохоос гадна +44с нөхцөлийн хэлбэр олон тоо (Fp), хамаатуулах "х" (Fh), үйл үгийн тодотгон холбох нөхцөл (Fd)-ийн зарим нөхцөлүүдтэй холбогдож болно гэж заажээ. Мөн нөхцөлийг үгийн үндсэнд залгахдаа үгийн

төгсгөлийн үсгийг тодорхойлох шаардлагатай. Иймээс ямар ямар үсгээр төгссөн үгэнд нөхцөлийн хэлбэрүүдийг залгаж болохыг харуулав. Үг хэрхэн задалсныг жишээгээр харуулав (хүснэгт-5).

ДУГНЭЛТ

1. Ц.Дамдинсүрэн, Б.Осор нар монгол хэлний толь хийхдээ үгийг кирил үсгийн зөв бичих дүрмийн дагуу ижил хувилах эсэхээр нь бүлэг болгон хуваасан байдаг. Ингэж үгийг бүлэглэж хуваах санаа бол уг толийг ашиглахад авсаархан болгох үүднээс юм. Харин цахим боловсруулалтад үгийн үндсийг ингэж бүлэг болгон хуваах шаардлагагүй. Харин нөхцөлийг кирил үсгийн зөв бичих дүрмийн дагуу ижил хувилах эсэхээр нь бүлэг болгон хуваах шаардлагатай байв.
2. Монгол хэлний нөхцөлүүд үгийн бүтцэд дараалан орохдоо өмнөх өмнөхөө дагасан бүтэцтэй байдаг. Өөрөөр хэлбэл нөхцөлийг залгаж болох эсэх нь өмнөх гишүүнээсээ хамааралтай байдаг. Иймээс нөхцөлийн хэлбэр бүрт залгагдах шинж үүсгээд ямар гишүүнтэй залгагдаж болохыг нөхцөлийнх нь шинжид зааж өглөө. Ингэснээр морфотактикийн сан үүсгэх шаардлагагүй болов. Мөн нөхцөлийн залгагдах боломжийг авиазүйн үүднээс тодорхойлсон тул хэлний хувьсал өөрчлөлт, зөв бичгийн дүрмийн алдаатай бичигдсэн үг зэргийг таних боломжтой болсон.
3. Хувилсан үгээс үгийн салаа утга ялгах нь чухал ач холбогдолтой. Учир нь кирил бичигт үг хувилсан хэлбэрээрээ тохиолдох нь элбэг (68%) байдаг. Энэ нь монгол хэлний залгамал бүтэцтэй холбоотой. Иймээс морфотактик ашиглан хувилсан үгээс үгийн салаа утгыг ялгав.
4. Орчин цагийн кирил бичигт зөв бичгийн дүрмийн алдаатай бичвэр их тохиолдох болсон. Иймээс зөв бичгийн алдаатай бичигдсэн үгийг бүтцээр задлахдаа үгийн санд үүсгэсэн морфотактикийн харьцуулах замаар зөв бичгийн дүрмийн алгааг засав. Зөв бичгийн дүрмийн алдааг дүрэм ашиглан автоматаар засварлах ажил хараахан хийгдээгүй шинэ ажил юм

АШИГЛАСАН ХЭВЛЭЛИЙН ЖАГСААЛТ

1. Нямаа Д. Түгээмэл хэл зүйн зарим зарчмыг монгол хэлнээ туршин үзэх нь [D]. Улаанбаатар хот: МУИС, 2006
2. Монгол хэл бичиг 2-р боть [M]. Улаанбаатар хот: Монгол улсын төрийн хэлний зөвлөл, 2006
3. Мөнх-Учрал Э. Хөрвүүлэх програмд зориулсан монгол хэлний судалгаа [D]. Улаанбаатар хот: МУИС, 2010
4. Дамдинсүрэн Ц., Осор Б. Монгол үсгийн дүрмийн толь[M]. Улаанбаатар хот: Ардын Боловсролын Яамны хэвлэл, Д.Сүхбаатарын нэрэмжит хэвлэлийн комбинат, 1983
5. GB/T 26235-2010, Information technology - Mongolian word and expression marks for information processing [S]. Beijing: China Standard Press, 2010
6. Нэргүй Б., Корфф О., нар. Монгол хэлний хэл шинжлэлд хоёр төвшинт морфологийн аргын хэрэглээ. Улаанбаатар: 1996
7. Батзолбоо Б. Монгол хэлний үг хувиллын компьютер загвар ба бичвэрийг хөрвүүлэх систем [D]. Улаанбаатар хот: Шинжлэх Ухаан, Технологийн Их Сургууль, 2012

ANALYZING WORD AS STRUCTURE IN ACORDANCE WITH POSSIBILITY OF ATTACHING INFLECTIONAL SUFFIX

ABSTRACT

There is a need to determine grammatical meaning emerged in word to conduct survey on Mongolian sentence structure. Saving every word inflected by grammatical affixes was unsuitable in terms of agglutinative Mongolian language. Thus, we aimed to separate lexical and grammatical meaning when analyze word as structure. However, misspelled words were being encountered in modern Cyrillic script. As a result, we need to separate lexical and grammatical meaning from misspelled scripts. Moreover, we created method analyzing word as structure due to nature of attaching inflectional suffix. We can separate lexical and grammatical meaning from misspelled words using the method. In addition, the method can recognize if it is noun or verb due to possibility of attaching inflectional suffix. Since noun can be not only inflected in different ways but also express different meaning because of inflectional suffix, the method cannot recognize. Thus, we used morphotactic to separate ambiguous meaning and fix error in spelling. It is important to separate ambiguous meaning from inflected words since word are commonly occurred (68%) as inflected form. In addition, it is new work as the work which corrects automatically error of spelling rule using rule has not been done yet. As you can see current e-processing in Cyrillic script, there is need to develop model inflecting words geared toward Mongolian language nature. Thus, we aimed to reflect grammar of Cyrillic script, phonetic principle and nature of Mongolian peoples' thinking on model inflecting word.

KEYWORDS: orphotactic, error in spelling rule, ambiguity