

Компьютерын ухаан

# Монгол хэлний сурах бичгийн эхийн тоон шинжилгээ: V анги

Н.Оюундарь<sup>1</sup>, М.Золжаргал<sup>2</sup>, Ч.Алтангэрэл<sup>2,\*</sup>

<sup>1</sup>МУИС, ШУС, ХУС, Монгол хэл, хэл шинжлэлийн тэнхим,

<sup>2</sup>МУИС, ХШУИС, Мэдээлэл, компьютерын ухааны тэнхим

Received on 2022.04.29; Revised on 2022.10.14; Accepted on 2022.11.29

\*Холбоо баригч зохиогч: altangerel@num.edu.mn.

## Хураангуй

Энэхүү судалгааны ажлын хүрээнд бид машин сургалтын аргад суурилан бүтээсэн монгол хэл боловсруулалтын хэрэглүүрүүд болон нээлттэй эхийн програмуудыг ашиглан монгол хэлний 5 дугаар ангийн сурах бичгийн эхийг боловсруулан шинжилж, үеийн давтамж, үгийн сангийн давтамж болон үгийн аймгийн хэрэглээ, хэл зүйн хувилал, өгүүлбэрийн бүтэц, хэв шинжийн давтамж зэрэг тоон шинжилгээ хийж, үр дүнг танилцуулахыг зорьсон болно. Сурах бичгийн эхийн шинжилгээ нь түүний агуулга, стандарт, сургалтын хөтөлбөрийн удирдамж боловсруулах болон анги дэвших бүрд агуулга, үгийн сан, хэл найруулгын талаас ямар ахиц дэвшил гарч байгааг тогтоох зэрэгт чухал ач холбогдолтой юм. Мөн цахим хэл шинжлэлийн үүднээс монгол хэлний боловсруулалтыг улам боловсронгуй болгоход нэмэр болно гэж үзэж байна. Монгол хэлний V ангийн сурах бичиг нь үлгэр, өгүүлэг, шүлэг, түүвэр өгүүлбэр, шинжлэх ухааны хялбаршуулсан нийтлэл, захидал, мэдээ, зохион бичлэг, илтгэл, өдрийн тэмдэглэл, жүжиг зэрэг 29,814 токен бүхий 3,082 өгүүлбэртэй 11 төрлийн 61 эхээс бүтсэн байна.

**Түлхүүр үг:** цахим хэл шинжлэл, хэл боловсруулалт, сурах бичиг, өгөгдлийн шинжилгээ, үгийн сан, хэл зүйн хувилал

## 1 Удиртгал

Хүүхэд эх хэлээ шинжлэх ухааны үүднээс сурч эзэмших анхны хэрэглэгдэхүүн нь тухайн хэлний сурах бичиг байдаг. Гийм учраас сурах бичгийг бодит үндэслэл, туршилт судалгаанд суурилан бүтээх ёстой юм. Улс орон бүрд тухайн орны хэлний онцлогт тохирсон стандартын баримт бичгийг боловсруулан ашигладаг. Жишээлбэл, бичвэрийн уншигдцыг тооцохдоо өгүүлбэрийн үг, үеийн тоо, урт зэргийг тооцож тогтоодог байна [1]. Монгол улсын хувьд 2008 онд батлагдсан “Бага, дунд боловсролын хэвлэмэл сурах бичигт тавих шаардлага” стандарт [2]-д дараах шаардлагуудыг тодорхойлсон байдаг. Үүнд:

**Ерөнхий шаардлага:** Суралцагчдын нас сэтгэхүйн онцлог, танин мэдэхүйн түвшинд нийцсэн байна.

**Агуулгын шаардлага:** Агуулга нь танин мэдэхүйн хөгжлийн үе шат, нас сэтгэхүйн онцлогт тохирсон, боловсролын түвшин, анги, бүлэг, сэдэв хоорондын залгамж холбоог хангасан байна.

**Хэл найруулгын шаардлага:** Танин мэдэхүйн тухайн түвшинд тохирсон хэл найруулгын болон зөв бичих дүрмийн алдаагүй, найруул-

га нь зөв оновчтой, товч тодорхой, энгийн ойлгомжтой, эмх цэгцтэй байна

гэж заажээ. Сурах бичгийн агуулга, хүртээмж, чанар, арга зүй зэргийг сурах бичгийн хэрэглэгчид тулгуурлан хийсэн судалгааны ажлууд цөөнгүй бий. Н.Бэгз, Б.Туяа нар [3] 5, 10 дугаар ангийн зарим сурах бичгийн агуулга, арга зүй, хэл найруулга, хэвлэлийн дизайнд багш нараас санал асуулга авах аргаар зарим шинжилгээ хийсэн бол Батбямба, Т. Алтантуяа, С. Алтангэрэл, Ч. нар [4] ЕБС-ийн I ангийн монгол хэлний сурах бичгийн үгийн сан, үгийн хувиллыг цахим хэл шинжлэлийн үүднээс бүртгэн судалсан байна. Мөн сурагчийн үгийн сан, өгүүлбэр бүтээх чадамжийг тодорхойлох судалгааны ажлууд ч хийгдсэн байна [5]. Дээр дурдсан сурах бичгийн ерөнхий болон агуулга, хэл найруулгын шаардлагыг хэрхэн хангаж байгаа эсэхийг тодорхойлоход сурах бичгийн эхийн өгөгдөлд боловсруулалт хийж судлах зайлшгүй шаардлагатай. Энэхүү ажилд монгол хэл боловсруулалтын олон чиглэлээр хөгжиж байгаа хөмрөгт суурилсан судалгааны ажлуудаас зөвхөн сурах бичгийн эхтэй холбоотой судалгааны ажлуудыг авч үзсэн болно.

## 2 Материал, Арга зүй

Судалгаанд ашигласан материал нь Олзвой Л, Оюунцэцэг Б, Энхжаргал М, Оюунцэцэг Ш, Монгол хэл V. Уб., 2020 (долоо дахь хэвлэл) [6] 152 хэвлэлийн хуудас бүхий сурах бичиг болно.

Судалгааны материалыг боловсруулсан дараалал:

- (1) PDF өргөтгөлөөс txt өргөтгөл болгон засаж, өгүүлбэрээр салгасан өгөгдөл (гарчиг, жишээ, эхэд холбогдоогүй холбоо үг, заавар, даалгаврын өгөгдлийг хассан)
- (2) Үгийн аймгийн тэмдэглүүр болон Өгүүлбэрзүйн задлуур программаар хагас автомат тэмдэглэгээ хийж өгөгдөл бэлдэх
- (3) Хагас автомат тэмдэглэгээт өгүүлбэрийг шалгаж гараар зассан өгөгдөл бэлдэх
- (4) Тоон шинжилгээг хийх.

Ингэхдээ Машин оюуны лабораторийн хөгжүүлсэн үеэр таслагч, монгол хэлний үгийн аймгийн тэмдэглүүр, нэрлэсэн нэгж таниур, өгүүлбэр зүйн задлуур [7] болон хөмрөг тэмдэглэх, шинжлэх *treetrans* [8], *tregex* [9], *antconc* [10] зэрэг нээлттэй эхийн програмыг ашиглаж, тоон шинжилгээний өгөгдлүүдээ гаргаж авав.

## 3 Үр дүн

Монгол хэлний V ангийн сурах бичиг нь үлгэр, өгүүлэг, шүлэг, түүвэр өгүүлбэр, шинжлэх ухааны хялбаршуулсан нийтлэл, захидал, мэдээ, зохион бичлэг, илтгэл, өдрийн тэмдэглэл, жүжиг зэрэг 29,814 токен бүхий 3,082 өгүүлбэртэй 11 төрлийн 61 эхээс бүтсэн байна.

Эдгээр эх дээр хэл боловсруулалтын хэрэглүүр [7] болон нээлттэй эхийн программуудыг ашиглан үгийн сан, үгийн үеийн давтамж, үгийн аймаг, үгийн хувилал, нэрлэсэн нэгж, өгүүлбэрийн бүтэц, хэв шинжийн түвшинд хийсэн тоон шинжилгээг дараах дэд бүлгүүдэд танилцуулав.

### 3.1 Үгийн сан

Сурах бичгийн 29,814 токен нь 6,543 дахин давтагдахгүй нэгжээс бүтсэн бөгөөд 100-аас олон удаа давтагдаж буй үг, бүтээврийг жагсаавал дараах байдалтай байна.

Сурах бичгийн үгийн сангийн давтамжийг тодорхойлж, анги тус бүрээр харьцуулснаар анги ахихад гарах шинэ үгийн хэрэглээг тодорхойлох боломжтой. Энэ арга зүйг ашиглан бусад олон хэлний адилаар монгол хэлний үгийн сангийн түвшинг ялган тодорхойлж, үгийн сангийн жагсаалт гаргаж болох юм.

Хүснэгт 1: Үгийн давтамж (100-аас дээш)

Төрөл	Эрэмбэ	Давтамж
<i>нь</i>	1	528
<i>гээж</i>	2	202
<i>байна</i>	3	184
<i>ямар</i>	4	176
<i>вэ</i>	5	172
<i>юм</i>	6	136
<i>хоёр</i>	7	127
<i>бэ</i>	8	120
<i>л</i>	9	116
<i>үг</i>	10	110
<i>нэг</i>	11	109
<i>бичээрэй</i>	12	104

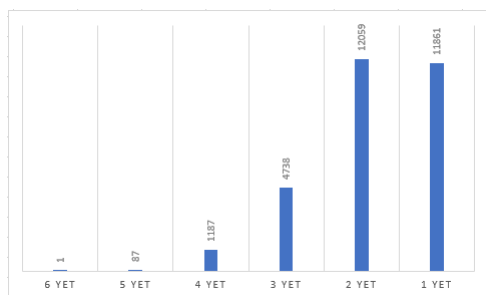
### 3.2 Үгийн үе

Үеэр таслах хэрэглүүр ашиглан нэгж үгийн үеийн тоо, өгүүлбэр дэх нийт үеийн тоо, нийт эх дэх үеийн давтамжийг тооцоолов. Сурах бичгийн эхийн үг нь 1-6 үетэй, өгүүлбэр нь 1-107 үетэй, нэг үе нь 1-6 хүртэл үсгээс бүтсэн байна. Өгүүлбэр дэх үеийн тоо, уртыг тодорхойлох нь сурах бичигт сонгон оруулж буй эхийн уншигдах байдал (reading ability)-ыг тодорхойлох, цаашилбал сурагчийн насны онцлог, танин мэдэхүйн чадварт тохируулан үеийн тоо, үгийн тоо, өгүүлбэрийн уртыг сонгох, анги дэвших бүрд ашиглах сурах бичгийн эхийн хүндрэлийн зэргийг тодорхойлох боломжийг олгох юм. Монгол хэлний V ангийн сурах бичгийн хувьд

- Нэг өгүүлбэрийн үеийн дундаж тоо 40.
- Хамгийн богино үг нь хоёр үсгээс бүтсэн би, та гэх мэт үг.
- Хамгийн олон үетэй үг нь 6 үетэй буюу “*төлөв-лө-гөө-ний-хөө*”
- Хамгийн олон үетэй өгүүлбэр нь “*Энэхүү сурах бичиг нь “Монгол Улсын Зохиогчийн эрх болон түүнд хамаарах эрхийн тухай” хуулиар хамгаалагдсан бөгөөд Боловсрол, Соёл, Шинжлэх Ухаан, Спортын Яамнаас бичгээр авсан зөвшөөрлөөс бусад тохиолдолд цахим болон хэвлэмэл хэлбэрээр, бүтнээр эсхүл хэсэгчлэн хувилал, хэвлэх, аливаа хэлбэрээр мэдээллийн санд оруулахыг хориглоно*”
- Хамгийн богино өгүүлбэр нь *Би.* гэсэн өгүүлбэр байна.

Хоёр үетэй үг нийт үгийн 40.3 хувийг, нэг үетэй үг 39.6 хувийг, хоёроос дээш үетэй үг 20.1 хувийг эзэлж байна. Энэ бичвэр дэх давтамж нь Ч.Алтангэрэл нарын [11] ажилтай харьцуулахад адилхан тархалттай нь харагдаж байна.

Э - эгшиг Г - гийгүүлэгч Өгөгдөлд ГЭГ гэсэн бүтэцтэй үе хамгийн олон давтамжтай орсон бол гурван Г дараалан орсон 5-6 үсгээр бүтсэн үе хамгийн цөөн байна.



Зураг 1: Үгийн үеийн давтамж

Хүснэгт 2: Үеийн давтамж

Төрөл	Эрэмбэ	Давтамж	Жишээ
гэг	1	16,888	гар
гэ	2	13,027	ха-вар
гээ	3	5,646	лаа
гээг	4	5,028	хуур
эг	5	2,457	ал-га
э	6	2,424	э-вэр
гэгг	7	1,881	багц
г	8	852	ч, л
эгг	9	597	эгч
ээ	10	372	аа-руул
гээгг	11	362	байрт
ээг	12	263	аар-хал
ээгг	13	62	аарц
ггэг	14	47	мян-ган
гэггг	15	2	хөрст
гээггг	16	2	гуурст
ээггг	17	1	үйлст

### 3.3 Үгийн аймаг

Машин сургалтын аргаар сургасан үгийн аймгийн тэмдэглүүр [12]-ээр автомат тэмдэглэгээ хийсний дараа гараар засаж, дараах өгөгдлийг боловсруулан гаргалаа. Монгол хэлний биед хамаатуулах минь, чинь тань, нь болон олон тооны нар, чиглэхийн тийн ялгалын руу рүү бүтээвэр, цэг тэмдэг зэрэг нь бие даасан нэгж болж үгийн аймгийн ангилалд орохгүй боловч тусдаа бичигдэж, танигддаг бичвэрийн бүрэлдэхүүн хэсэг учраас тус хүснэгтэд багтаан шинжиллээ.

Сурах бичгийн эхийн нийт өгөгдлийн 35.5 хувийг жинхэнэ нэр, 23.7 хувийг үйл үг, 14.56 хувийг цэг тэмдэг эзэлсэн байна. Мөн 6.73 хувь нь жинхэнэ нэрийн хэлбэр дүрс, шинж чанарыг заах үүрэгтэй тэмдэг нэр, 5.11 хувь нь жинхэнэ нэрийг төлөөлөн заах үүрэгтэй төлөөний үг байна.

### 3.4 Үгийн хувилал

Монгол хэлний нэр үг нь үгийн үндэс, олон тоо, тийн ялгал, хамаатуулах нөхцөл гэсэн дарааллаар хувилдаг байдаг бол үйл үг нь үндэс, хэв, байдал, холбох төгсгөх нөхцөл гэсэн дарааллаар хувил-

Хүснэгт 3: Үгийн аймгийн давтамж

Үгийн аймаг	Давтамж	Жишээ	Хувь
Жинхэнэ нэр	10,587	аав	35.51
Үйл үг	7,073	яв-	23.72
Цэг тэмдэг	4,341	. , "	14.56
Тэмдэг нэр	2,005	улаан	6.73
Төлөөний үг	1,524	та чи	5.11
Орон цагийн нэр	814	дээр	2.73
Бусад	798	нар, руу	2.68
Холбох үг	776	харин	2.60
Хамаатуулах	644	нь, минь	2.16
Тооны нэр	524	гурав	1.76
Баймж	395	ёстой	1.32
Аялга үг	94	аа	0.32
Туслах үйл үг	90	бай- бол-	0.30
Үгүйсгэх	62	биш	0.21
нэр үг+гүй	60	наргүй	0.20
Товчилсон үг	27	МУИС	0.09

даг. Өгөгдлийн хувьд хамгийн идэвхтэй хэрэглэгдсэн тийн ялгал нь нэрлэхийн тийн ялгал байна. Энэ нь нэр үг өгүүлбэрийн эзэн биеийн үүргээр хамгийн олон удаа орсон гэсэн үг юм. Хамгийн бага хэрэглэгдсэн нь чиглэхийн тийн ялгал бөгөөд орон зай, чиглэл заах байц гишүүний үүргээр орсон нэр үг хамгийн цөөн байгааг илтгэнэ. Нэр үгийн хувиллын давтамжийг Хүснэгт 6, үйл үгийн хувиллын давтамжийг Хүснэгт 7-д жишээний хамт үзүүлээ.

Өгөгдлийн нийт нэр үгийн 58.7 хувь нь нэрлэхийн тийн ялгал (багш), 12.8 хувь нь харьяалахын тийн ялгал (багшийн), 15.5 хувь нь заахын тийн ялгал (багшийг), 5.4 хувь нь өгөх оршихын тийн ялгал (багшид), 1.9 хувь нь гарахын тийн ялгал (багшаас), 3 хувь нь хамтрахын тийн ялгал (багштай), 2.4 хувь нь үйлдэхийн тийн ялгал (багшаар), 0.2 хувь нь чиглэхийн тийн ялгал (багш руу)-аар хэлбэржсэн байна. Жинхэнэ нэрийг төлөөлөн нэршиж, дутмаг хувилдаг тэмдэг нэр, тооны нэрийн хамгийн идэвхтэй хэрэглэгддэг тийн ялгал нь тэмдэг нэрийн хувьд үйлдэхийн тийн ялгал (улаанаар), тооны нэрийн хувьд өгөх орших (хоёрт)-ын тийн ялгал байна. Үйл үгийн холбох, төгсгөх нөхцөлийн хэрэглээг тодорхойлж үзэхэд нийт үйл үгийн 36.5 хувь нь өөрөө үйлдэх хэв, энгийн үйлдэх байдалтай тодотгон холбох нөхцөл (яв+Ө+Ө+сан)-өөр, 30 хувь нь нөхцөлдүүлэн холбох нөхцөл (яв+Ө+Ө+эс)-өөр хэлбэржсэн байна. Биеэр төгсгөх нөхцөлөөр 10 хувь нь хэлбэржжээ. Үйл үгийн бүтээврийн нийлмэл хэлбэрүүд сурах бичгийн эхэд түгээмэл бус ашиглагддаг нь харагдаж байна.

### 3.5 Нэрлэсэн нэгж

Хэл боловсруулалтын олон зорилгын нэг нь мэдээлэл задлах байдаг бол мэдээлэл задлах системийн хамгийн чухал даалгавар нь нэрлэсэн нэгж таних явдал. Өмнөх судалгаануудаар нэрлэсэн нэг-

Хүснэгт 4: Нэр үгийн хувиллын давтамж

Хувилал	Нэрлэх	Харьяалах	Заах	Өгөх орших	Гарах	Хамтрах	Чиглэх	Үйлдэх	Нийт
Ерийн нэр	3530	1529	1966	662	244	383	26	302	8642
Ерийн нэр олон тоо	1358	39	20	14	2	2			1435
Оноосон нэр	375	85	28	14	4	4			510
Тэмдэг нэр	1957	5	10	6	1	8		18	2005
Тооны нэр	478	15	14	16		1			524
Нийт	7698	1673	2038	712	251	398	26	320	13116
Эзлэх хувь	58.7	12.8	15.5	5.4	1.9	3	0.2	2.4	

Хүснэгт 5: Үйл үгийн хувиллын давтамж

Хэв	Байдал	Нөхцөлдүүлэн холбох	Тодотгон холбох	Өнгөрсөн цаг	Одоо цаг	Биеэр төгсгөх	Нийт	Эзлэх хувь	Жишгээ
							2,585	36.55	яв-сан
							2,176	30.76	явж
							749	10.59	явъя
							507	7.17	явна
							496	7.01	явуулья
							431	6.09	явжээ
							29	0.41	явуулаад
							26	0.37	явуулсан
							13	0.18	явчихжээ
							13	0.18	явчхаад
							13	0.18	явчихлаа
							12	0.17	явчихсан
							8	0.11	явууллаа
							6	0.08	явчихъя
							5	0.07	явуулна

жид хүн, байгууллага, газар орны нэр, огноо болон цаг хугацаа, мөнгөн тэмдэгт болон тоог багтааж иржээ [13] [14]. Сурах бичгийн агуулгад нэрлэсэн нэгж таних нь дээр дурдсан стандарт заасан "ухагдахуун, хууль, зүй тогтлыг хэрэглээ болон танин мэдэхүйн түвшинд тохируулан буулгаж, зөв оновчтой томьёолсон, шинжлэх ухааны хувьд алдаагүй, Монгол улсын хөгжлийн болон даяаршлын тулгамдсан асуудлыг тусгасан байх [2] гэсэн шаардлагыг шалгах нэг үзүүлэлт болж болох юм.

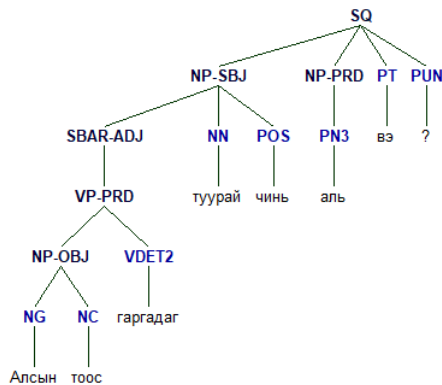
Хүснэгт 6: Нэрлэсэн нэгж

Төрөл	Тоо	Дундаж	Жишээ
Хүн (PER)	104	1.7	Цолмон
Газар орон (LOC)	79	1.1	Монгол улс
Байгууллага (ORG)	50	1.4	Монголын урлагийн зөвлөл
Огноо (DATE)	5	2.2	2015 он
Нийт	238	1.5	

Нэрлэсэн нэгжийн таниур ашиглахад сурах бичгийн эхээс нийт 238 нэрлэсэн нэгж олдсон бөгөөд үүнээс 104 нь хүний нэр, 79 нь байр орны нэр, 50 нь байгууллагын нэр, 5 нь огноо заасан байна. Хүний нэр дунджаар 2 үгтэй, газар орны нэр 1 үгтэй байна.

### 3.6 Өгүүлбэрийн бүтэц, хэв шинж, давтамж

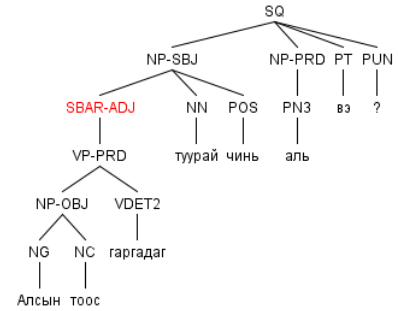
Монгол хэлний V ангийн сурах бичгийн эхийн үгийн аймгийн тэмдэглүүрт [12] [15] өгөгдлийг өгүүлбэр зүйн задлуур программ ашиглах мөрдлөгийн [16] [17] дагуу хагас автоматаар боловсруулан, гараар засахад сурах бичгийн 61 эх 3082 өгүүлбэрээс бүтсэн байна. Өгүүлбэр зүйн задлуур программ ашиглан



Зураг 2: Treetrans программаар үгийн аймаг, өгүүлбэрийн тэмдэглэгээ хадсан байдал

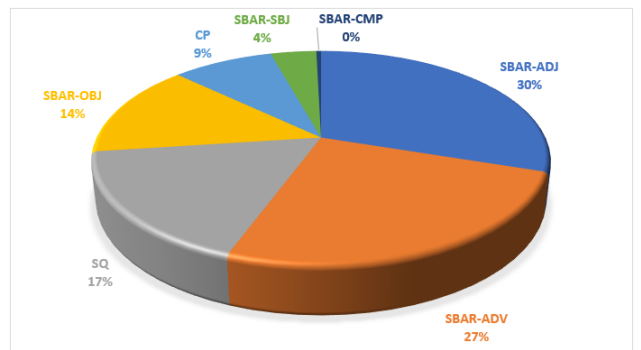
үгийн аймгийн болон нийлэмж, өгүүлбэр зүйн тэмдэглэгээ хадсан өгөгдлөөс нээлттэй эхийн tregex [9]

программ ашиглан тодотгол, тусагдахуун, байц гишүүн өгүүлбэр, оруулсан өгүүлбэр, асуух өгүүлбэр гэх мэт өгүүлбэрийн нэгж хэсгүүдийг ялган авч, шинжиллээ. Нийт өгөгдөлд нэг бүрэлдэхүүнт эзэн



Зураг 3: Tregex программаас тодотгол гишүүн өгүүлбэрийг ялган харуулах явц

биегүй өгүүлбэр буюу зөвхөн өгүүлэхүүний бүлгээр илэрсэн өгүүлбэр 1289, гарчиг болон бусад өгүүлбэр биш бүтэц 559 байсан бол хамгийн түгээмэл хэлбэр болох хоёр болон гурван бүрэлдэхүүнтэй энгийн, нийлмэл өгүүлбэрүүдээс бусад асуух, оруулбар, гишүүн өгүүлбэрийг бүртгэн харьцуулбал дараах байдалтай байна.



Зураг 4: Гишүүн өгүүлбэрийн харьцаа

Эдгээр үзүүлэлтээс монгол хэлний гишүүн өгүүлбэрийн бүтэц, харьцаа, давтамж, утга, үүрэг цаашлаад гол өгүүлбэрийн бусад гишүүнтэй холбогдох арга, хэрэглүүрийг тодорхойлж судлах боломжтой.

Өгөгдлийн нийт гишүүн өгүүлбэрийн 30 хувийг нэр үгийг нэмэн дэлгэрүүлдэг тодотгол гишүүн өгүүлбэр, 27 хувийг үйлийн орон байр, цаг хугацаа, болзол шалтгааныг илэрхийлэх байц гишүүн өгүүлбэр, 14 хувийг шууд тусагдахуун гишүүн өгүүлбэр эзэлж байна. Хүснэгт 5-д гишүүн өгүүлбэрийн давтамжийг жишээний хамт үзүүллээ.

## 4 Хэлэлцүүлэг

Сурах бичгийн агуулга, арга зүй, орчин, хэрэглэгчид тулгуурласан олон төрлийн судалгаа бий. Гэвч

## Хүснэгт 7: Өгүүлбэрийн төрөл

№	Гишүүн өгүүлбэр	Тэмдэглэгээ	Тоо
1	Тодотгол гишүүн өгүүлбэр “Гэтэл {тэр үүрийг анх олж харсан} бялзуухай багтсангүй”	SBAR-ADJ	552
2	Байц гишүүн өгүүлбэр {Шуурга ихтэй учраас} чоно эргээд байдаг юм шүү	SBAR-ADV	466
3	Асуух өгүүлбэр Хэн маань идэх вэ?	SQ	316
4	Шууд тусагдахуун гишүүн өгүүлбэр {Нарангийн бичсэнийг} ажиглаарай.	SBAR-OBJ	260
5	Хөндлөнгийн оруулбар өгүүлбэр “Би толио арчиж байна” гэв.	CP	158
6	Өгүүлэгдэхүүн гишүүн өгүүлбэр {Гэрийн хойморт аав суух нь} сайхан.	SBAR-SBJ	71
7	Шууд бус тусагдахуун гишүүн өгүүлбэр Өөрийнхөө бодлыг {бусдын ярьсантай} харьцуулан сонсоно.	SBAR-CMP	8

анги дэвших бүрд сурах бичгийн агуулга хэрхэн хүндэрч байгааг өгөгдөлд суурилан бодитой тодорхойлох шаардлага байсаар байна. Ялангуяа бага ангийн монгол хэлний сурах бичгийн хувьд үгийн сан болон хэл зүйн түвшнийг тодорхойлж, үгийн уншигдах байдал (readability)-г тогтоох зайлшгүй шаардлагатай байна. Энэ нь сурах бичиг, хэлний хэрэглэгдэхүүн зохиох, цаашлаад монгол хэлний түвшин тогтоох шалгалтын стандарт, хэрэглэгдэхүүн боловсруулахад чухал ач холбогдолтой юм.

## 5 Дүгнэлт

Бид энэхүү судалгаагаар Машин оюуны лабораторийн судлаачдын багийн боловсруулсан үеэр таслагч, монгол хэлний үгийн аймгийн тэмдэглүүр, нэрлэсэн нэгж таниур, өгүүлбэр зүйн задлуур программ болон *tretrans* [8], *tregex* [9], *antconc* [10] зэрэг нээлттэй эхийн программ ашиглан монгол хэлний V ангийн сурах бичгийн эхийг үе, үгийн сан, үгийн аймаг, хэл зүй, өгүүлбэр зүйн түвшинд бүртгэн тэмдэглэж, тоон шинжилгээ хийлээ. Сурах бичгийн 29814 токен бүхий үлгэр, шүлэг, зохион бичлэг, өдрийн тэмдэглэл зэрэг 11 төрлийн 61 эх нь 3082 өгүүлбэрээс бүтсэн бол үйл үгийн 36 хувь нь тодотгон холбох нөхцөлөөр, нэр үгийн 58.7 хувь нь нэрлэхийн тийн ялгалаар хэлбэржжээ. Бидний судалгаа монгол хэлний боловсролын талаас хүүхдийн нас, сэтгэхүй, танин мэдэхүйн онцлогт тохирсон сурах бичиг, хэрэглэгдэхүүн зохиох, монгол хэлний мэдлэгийн түвшин, хэм хэмжээг олон хэлний түвшин тогтоох шалгалтын арга аргачлалаар тодорхойлох, цахим хэл шинжлэлийн үүднээс хэл боловсруулалтын арга, алгоритм ашиглан үгээр салгах, өгүүлбэрээр салгах, үг зүйн болон өгүүлбэр зүйн бүрэн тэмдэглэгээ хадах, нэрлэсэн нэгж таних, цаашлаад монгол бичигт хөрвүүлэх, үгийн утга ялгах, ярин боловсруулах чиглэлээр олон талын судалгаа хийхэд нэг алхам болно гэж үзэж байна. Энэхүү ажлын су-

далгааны материалыг бусад ангийн сурах бичгийн эхээр өргөтгөн анги хоорондын ялгааг харьцуулан дүн шинжилгээ хийх, цаашлаад бусад хичээлийн сурах бичгийн эхтэй харьцуулан үзвэл зохистой байна.

## Талархал

МУИС-ийн Машин оюуны лабораторийн судлаачид болон “Монгол хэлний стандарт шалгалт” төслийн удирдагч доктор Д.Заяабаатар, төслийн багийн хамт олонд талархал илэрхийлье.

## Зохиогчийн оролцоо

Энэхүү судалгааны ажлын цахим хэрэглүүрүүдийг Ч.Алтангэрэл, М.Золжаргал нар зохиож, хагас автомат өгөгдөл бэлдэн, Н.Оюундарь гараар засаж сайжруулсан. Ч.Алтангэрэл засаж сайжруулсан өгөгдлөөс үе, үгийн сан, үийн аймаг, хэлзүйн тэмдэглэгээг ялгаж, Н.Оюундарь давтамжийг тодорхойлж, тоон шинжилгээг гүйцэтгэв. Өгүүллийг хамтран бичив.

## Ашиг сонирхлын зөрчилгүйн баталгаа

Ашиг сонирхлын ямар нэгэн зөрчилгүй болно.

## Ашигласан ном

- [1] Zorbaz K. An Evaluation on the Word-Sentence Lengths and Readability Levels of Tales in Turkish Textbooks. *Journal of Theory and Practice in Education*. 2007 01;3.
- [2] *estandard.gov*. Бага дунд боловсролын хэвлэмэл сурах бичигт тавих шаардлага /[www.estandard.gov.mn](http://www.estandard.gov.mn). 2008.

- [3] Бэгз Н, Туяа У. Дунд сургуулийн v, x ангийн сурах бичигт хийсэн судалгааны зарим дүн ба сурах бичгийн хэл найруулга. ЕБС-ийн сурах бичгийн хэл найруулга” сэдэвт эрдэм шинжилгээний хурлын илтгэлийн эмхэтгэл. 2019;3-9 тал.
- [4] Батбаямба Т, Алтангуяа С, Алтангэрэл Ч. ЕБС-ийн I ангийн монгол хэлний сурах бичгийн үгийн сан, үгийн хувиллыг цахим хэл шинжлэлийн үүднээс бүртгэн судалсан нь. Хэрэглээний хэл шинжлэл. 2021;VI(21):89-103.
- [5] Дэлгэрцэцэг Г. Монгол хүүхдийн өгүүлбэр бүтээх чадамжийн судалгаа. Магистрын зэрэг горилсон бүтээл. 2017.
- [6] Олзвой Л, Оюунцэцэг Б, Энхжаргал Б, Оюунцэцэг Ш. Монгол хэл V. сурах бичиг. 2020;VII.
- [7] Монгол хэл боловсруулалтын үзүүлэн хуудас [shorturl.at/jBCFR](http://shorturl.at/jBCFR). Machine Intelligence Laboratory NUM. 2016.
- [8] Bird S, Maeda K, Ma X, Lee H, Randall B, Zayat S. TableTrans, MultiTrans, InterTrans and TreeTrans: Diverse Tools Built on the Annotation Graph Toolkit. 2002 05.
- [9] Levy R, Andrew G. Tregex and Tsurgeon: tools for querying and manipulating tree data structures. 5th International Conference on Language Resources and Evaluation (LREC 2006). 2006.
- [10] Anthony L. AntConc (Version 4.0.10) [Computer Software]. 2022.
- [11] Алтангэрэл Ч, Энхжаргал М. Цахим бичвэрт суурилсан монгол хэлний үеийн давтамжийн судалгаа. Монгол судлал эрдэм шинжилгээний бичиг. 2013;VII.
- [12] Золжаргал М. Гурвалсан дүрэмт үгийн аймгийн тэмдэглүүр. Магистрын зэрэг горилсон бүтээл. 2011.
- [13] Золжаргал М, Алтангэрэл Ч, Оюундарь Н, Нанзадрагчаа Д. Монгол хэлний нэрлэсэн нэгжийн хөмрөг байгуулах нь. Монголын мэдээллийн технологи-2015. 2015.
- [14] Золжаргал М. Монгол хэлний нэрлэсэн нэгж таниур. Докторын зэрэг горилсон нэг сэдэвт бүтээл. 2017.
- [15] Purev J, Odbayar C. Part of Speech Tagging for Mongolian Corpus. Proceeding ALR7 Proceedings of the 7th Workshop on Asian Language Resources. 2009;VII.
- [16] Purev J, Altangerel C. An outline of Mongolian syntax. Мэдээллийн Технологийн Сургуулийн эрдэм шинжилгээний бичиг. 2002;I.
- [17] Оюундарь Н. Өгүүлбэр зүйн задлуур программд зориулан монгол хэлний өгүүлбэрийн бүтцийг тэмдэглэх нь. Магистрын зэрэг горилсон бүтээл. 2016.