

**ЦАХИМ БИЧВЭРТ СУУРИЛСАН
МОНГОЛ ХЭЛНИЙ ҮЕИЙН ДАВТАМЖИЙН СУДАЛГАА**

Ч. Алтангэрэл, М. Энхжаргал *

Товч агуулга

Хэлийг цахимаар боловсруулахад зайлиггүй хэрэглэгдэх чухал зүйл бол тухайн хэлний мэдлэгийг шингээсэн үнэн бодит, их хэмжээний эхийн цахим сан (corpus) юм. Цахим санд тулгуурлан хэл шинжлэлийн олон сонирхолтой судалгаа, дүгнэлт хийж болох бөгөөд үүний нэг нь үеийн давтамжийн судалгаа юм. Энэ судалгааг цахимаар хийх нь ихээхэн цаг, хүч хөдөлмөр хэмнэсэн ажил бөгөөд богино хугацаанд, үнэн зөв, цэгцтэй мэдээлэл, үр дүн гаргадаг онцлогтой.

Уг өгүүлэл нь монгол хэлний кирил бичгийн цахим санд тулгуурлан үеийн давтамжийг судалсан байдал, үр дүн, түүний хэрэглээ ач холбогдлын тухай өгүүлнэ.

Түлхүүр үг

Корпус хэл шинжлэл, цахим хэл шинжлэл, монгол хэлний үе.

Оршил

Корпус хэл шинжлэл олон талын ач холбогдолтой. Ялангуяа хэл шинжлэлд цагаан толгойн үсгийн хэрэглэгдэх давтамж; үгсийн аймгийн өгүүлбэр дэх үүрэг, хэрэглээ; дагаврын чадамж, тархац; оносон хайлтаар мэдээлэл авах; төс утгатай үгийг ялгах; зөв бичих дүрмийн алдаа засах; толь бичгийн судалгаа гээд бүр цаашилбал машин орчуулгын судалгаанд үндэс суурь нь болж өгдөг.

Корпуст суурилсан хэл шинжлэлийн судалгааг дотор нь уран зохиолын, үгийн сангийн, нутгийн аялгууны, боловсролын, хэлзүйн гэж ангилдаг. Жишээ нь, уран зохиолын судалгааг Библийн үг хэллэгийн судалгаа, тодорхой нэг зохиогчийн (Жишээ нь Шекспир) уран бүтээлийн үг хэллэгийн онцлогийг судлах гэх мэт зорилгоор хийдэг. Жишээ нь, Библид *how, you, he, once, between* зэрэг үгс хамгийн их давтамжтай; *how long, how many, how much, how much less, how much more, how often* зэрэг хэллэгийг (collocations) хамт хэрэглэсэн тохиолдол хамгийн их байгааг тогтоожээ[Graeme Kennedy, 1998,13-17].

Корпуст суурилсан судалгааны гол зорилтуудын нэг нь сургалтын хэрэглэгдэхүүн бий болгох, тухайлбал, хоёрдогч хэлийг сурч байгаа хүнд зориулан тухайн хэлэнд

* МУИС Компьютер хэл шинжлэлийн судалгааны төв, www.crlp.num.edu.mn

хамгийн их хэрэглэгддэг үгийн давтамжийн жагсаалтыг гаргах зэрэг юм[М.Энхжаргал, Ж.Пүрэв, Д.Нанзадрагчаа, 2012].

Зарим корпусыг тухайн хэлний хэлзүйг судлах зорилгоор бүтээсэн байдаг. Тухайлбал, Фрайс хэлний хэрэглээний нийгмийн ангийн ялгааг тодорхойлох зорилгоор “Америкийн англи хэлний хэлзүй” (1940) номдоо боловсрол болон нийгмийн орчноороо ялгаатай хувь хүмүүсийн АНУ-ын Засгийн газарт бичсэн захиа бүхий корпусыг оруулан ашиглажээ. Түүний судалгаанаас олон сонирхолтой дүгнэлт гарснаас дурдвал, төгс өнгөрсөн цагийн *done* гэсэн үг боловсролтой хүний захидалд харьцангуй цөөн тохиолдож байжээ[Graeme Kennedy, 1998,17].

Дээр дурдсанчлан хэлний судалгаанд корпус суурилсан статистик судалгаа чухал байр суурь эзэлдэг билээ. Монгол хэлний уламжлалт судалгаанд үеийн бүтэц, хувьслыг судлах нь нэг талаар эртний монгол хэлний үг, морфемын бүтцийг түүхэн үүднээс сэргээн тогтоох, нөгөө талаар орчин цагийн монгол хэлний авианы байрлалын хуулийг нарийн тусгасан зөв дуудах нормыг тогтоож, зөв бичих дүрмийг боловсронгуй болгоход онол практикийн ач холбогдолтой байдаг бол [Д.Төмөртоогоо, 1986] корпус хэл шинжлэлийн үүднээс монгол хэлний үеийн судалгааг хийх дараах шаардлага тулгарсан юм.

Үүнд МУИС-ийн Мэдээллийн технологийн сургуулийн Компьютер хэл шинжлэлийн судалгааны төвд (Center for research on language processing) хэрэгжүүлж буй “Бичвэрийг ярианд хөрвүүлэх” (Text-To-Speech), “Англи-Монгол орчуулгын онлайн програм” зэрэг төслийн хүрээнд, монгол үгийг дуудахад тухайн үг ямар үе, хэсгээс бүтэж байгааг нарийвчлан тогтоох, үгийг үеэр болон үгийн бүтцээр задлах шаардлага гарч ирсэн юм.

Мөн түүнчлэн үсгийн болон үеийн давтамжийн судалгаа нь компьютерийн гарын үсгийн байрлалыг шинээр загварчлан тогтооход чухал. Тухайлбал, бай- гэсэн үе монгол хэлэнд их давтамжтай байвал уг үеийг бүрдүүлж байгаа *a*, *y* гэсэн хоёр үсгийг компьютерийн гарт ойрхон (нэг дор) байрлуулах жишээтэй. Иймэрхүү тохиолдол орос үсгийн гарын байрлалд байдаг. Жишээ нь: про, ыва, ить гэсэн үсгийн хослол орос хэлэнд их тохиолддогийг харгалзан уг 3 үсгийн хослолыг зэрэгцүүлэн байрлуулсан байна.

Монгол хэлний үеийн судалгаа

Амьсгалын нэг түрэлтээр хэлэгдэх нэг буюу хэдэн авиаг үе гэх бөгөөд *ман-дуулсан, хой-но, хойш-ло-ход, тэм-дэг-лэг-дэх-дээ, хонь, морь, о-ё* гэх мэтээр нэг мөрөөс нөгөө мөрт үгийг шилжүүлэхдээ үеэр тасалж шилжүүлнэ[Ц.Дамдинсүрэн, Б.Осор нар, 1983]. Авиан зүйн үүднээс монгол хэлний үеийг акад. Б.Я.Владимирцов CV, CVC, VC гэсэн 3 янзын бүтэцтэй [Владимирцов Б.Я., 1929]., проф Ш.Хаттори CV [Hattori Sh, 1951], Д.Амоголов, И.Д.Бураев, акад Ш.Лувсанвандан нар V, VC, CV, CVC гэсэн 4 янзын бүтэцтэй гэж тэмдэглэжээ [Амоголов Д.1958, Бураев И.Д.1987, Лувсанвандан Ш.2010]. Монгол үеийн түүхэн хувьслын талаар Удо Пош, П.Ц.Биткеев, Ж.Төмөрцэрэн, С.Галсан, Х.Курибаяши нар сонирхолтой судалгааг хийсэн байдаг [Төмөртоогоо Д, 1992]. Үеийн хувьслыг түүхэн үүднээс авч үзсэн судалгааг акад Д.Төмөртоогоо [Д.Төмөртоогоо, 1987,1992], проф М.Базаррагчаа [М.Базаррагчаа, 2002] нар хийж монгол хэлний үеийн тухай судалгааг улам баяжуулсаар байна. “Эртний монгол хэлэнд Э, ГЭ,

ЭГ, ГЭГ гэсэн үеийн 4 хэлбэр байх ба ГЭГГ гэсэн үе хааяа тохиолдоно[Санжаа Ж., Надмид Д, 2008]” гэжээ.

Харин хэлний бодит материал дээр тулгуурлан монгол хэлний кирил үсгийн үеийн хэлбэрийг судалсан судалгаа, тухайлбал монгол хэлний кирилл бичвэрт ямар ямар төрлийн үе тохиолдож байна, ямар төрлийн үе хамгийн их давтамжтай байна гэх мэтээр нарийвчилсан судалгаа “Орчин цагийн монгол хэлэнд байгаа үеийн Э, ГЭ, ЭГ, ГЭГ хэлбэрээс давтамжийн судалгаа хийхэд Э, ЭГ гэсэн үе арваад хувийг эзлэх бөгөөд ГЭГ үе зонхилох суурийг буюу 50 гаруй хувийг эзэлдэг...өнөөгийн монгол хэлэнд битүү үе задгай үеэсээ илүү олон тохиолдоно” гэсэн [Санжаа Ж., Надмид Д, 2008] гэж үзсэнийг эс тооцвол бараг байхгүй байна.

Энэ ажлын онцлог нь цахим санд суурилсан үеийн хэлбэрийг тооцоолон гаргахад оршино. Энд хэлний бодит материал буюу 5,317,834 сая гаруй үгээс бүрдэх эхийн цахим санд суурилан үеийн хэлбэрийн судалгааг нарийн тооцон гаргасан юм. Ингэхдээ үгийг үелэх зарчмын дагуу зааглан хуваагаад эгшиг, гийгүүлэгчээр нь солин автоматаар гаргалаа. Уг судалгаанд хэрэглэсэн энэхүү 5 сая гаран үгтэй кирил монгол хэлний эхийн санг хойшид судалгааны эх гэнэ. Судалгааны эхээс гарсан давтамжийн үр дүнг монгол хэлний бодит онцлогоор төлөөлүүлэн үзэж болно.

Монгол хэлний цахим санд тулгуурлан монгол хэлний үеийн давтамжийг судалсан нь

Судалгааны эхэд давхардаагүй тоогоор 170 000 орчим үг байгаа бөгөөд үүнд уран зохиолын эхээс роман, тууж, өгүүллэг; сонин нийтлэлийн эхээс мэдээ, тэргүүн өгүүлэл (Үнэн сонины), нийтлэл, ярилцлага, сурвалжлага, өгүүлэл; хуулийн эхээс үндсэн хууль болон бусад хуулийг сонгон авсан болно. Хугацааны хувьд эдгээр бичвэр нь 1920-2008 онд хамаарч байгаа бөгөөд гадаадын болон эртний уран зохиолоос бичвэр ороогүй.

AntConc¹ програмыг ашигласан тооцооллын дүнд, судалгааны эхэд 18,088 төрлийн нийт 11,123,714 үе байна. Үүнээс үзэхэд монгол хэлний нэг үг дунджаар 2 үеэс бүтэж байна. Хүснэгт 1-ээс монгол хэлэнд хамгийн их давтамжтай үеийн жагсаалтыг жишээтэй хамт харууллаа.

#	Давтамж	Үе	Жишээ нь
1	178375	бай	*нийг мийн *дэвш лийн *тө лөө *өр нүүлж *бай гаа *улс *түм ний *
2	143260	а	*бай гуул ла гыг *до тоод *а жил *, *аж *а хуйн *бай гуу лал тыг
3	140774	га	үр *нө лөө тэй *ар га *хэм жээ *авч *бай вал *зо хи но *
4	108008	сан	*хамт *ол ны *өм нө *тул гар сан *зо рил тыг *ү зэл *сурт лын
5	105830	на	*өм нө *зог сож *бай гаа *юм *шиг *са наг да на
6	105509	нь	*төс лийг *байн гын *хо роо дод *харь яал лаар *нь *ху ваа ри лах ;
7	102691	нэ	*бай гуул лын *а жилд *ши нэ *ша танд *гар гах *шаард ла га *дээр хээс
8	101096	гүй	бү рэн *дүү рэн *а шиг лаж *ча дах гүй *бай на *. энд
9	97765	о	*хам гаа лах *, *нам *, *о лон *ний тийн *бай гуул ла га *
10	91318	ны	*эх лэл *хийн *яа мын *ан хан *шат ны *бай гуул ла гуу дын *дай чин

Хүснэгт 1. Судалгааны эх дэх хамгийн их давтамжтай үеийн жагсаалт

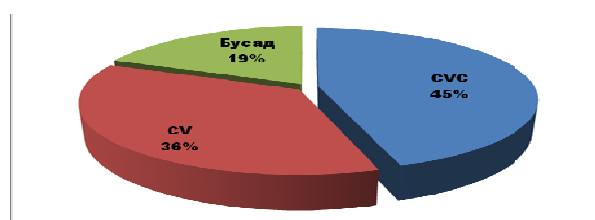
¹) Уг програмын AntConc3.2.4 хувилбарыг <http://www.antlab.sci.waseda.ac.jp/software.html> хаягаас татаж авч болно.

Энэхүү судалгааны эхэд 7 төрлийн загвартай үе тохиолдож байна. Уг загварыг гаргахдаа гадаад үг, цэг тэмдэг, дэс дугаар, товчилсон үг, бүтээвэр зэргийг (3500 орчим) оруулаагүй болно (Хүснэгт 2-г үзнэ үү).

#	Давтамж	Үеийн загвар	Жишээ
1	5,006,222	CVC	час, хүн, хар
2	3,924,647	CV	бай, га, на
3	624,101	CVCC	бөгж, боомт, дэрс
4	576,856	V	а, о, у
5	572,246	VC	өөс, үүх, эх
6	159,688	VCC	эмх, үйлд, эрх
7	3,217	CVCCC	авст, эрхт

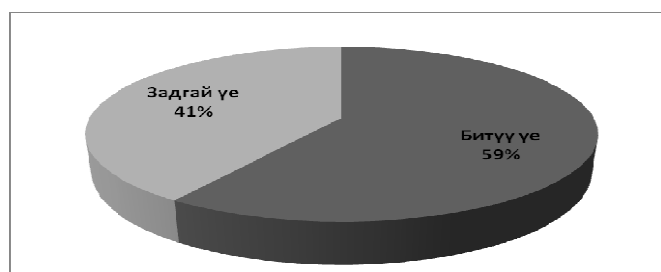
Хүснэгт 2. Монгол хэлний материалын санд тохиолдож буй үе

Хамгийн их давтамжтай үе нь CVC байна (Зураг 2-г үзнэ үү). Энэ нь “Санжаа Ж, Надмид Д (2008) нарын CVC гэсэн үе монгол хэлэнд зонхилох суурь буюу 50 гаруй хувийг эзэлж байна” гэсэн судалгаатай таарч байгаа юм.

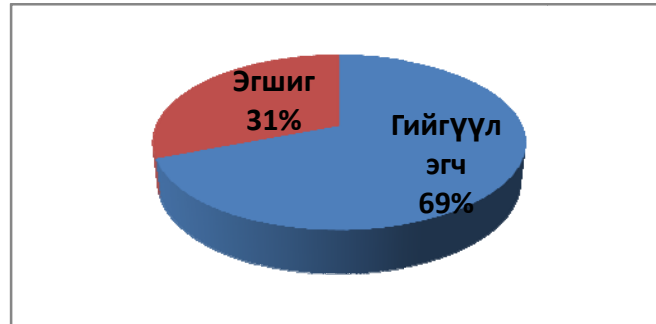


Зураг 2. Монгол хэлний үеийн төрлийн эзлэх хувь

Цаашилбал, битүү үе буюу гийгүүлэгчээр төгссөн үе болон гийгүүлэгч үсгийн хэрэглээ эгшгээс их байна (Зураг 3, 4-г үзнэ үү).



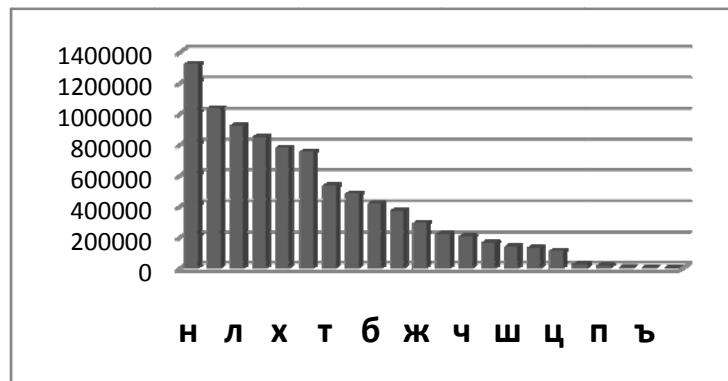
Зураг 3. Монгол хэлний битүү болон задгай үеийн хэрэглээ



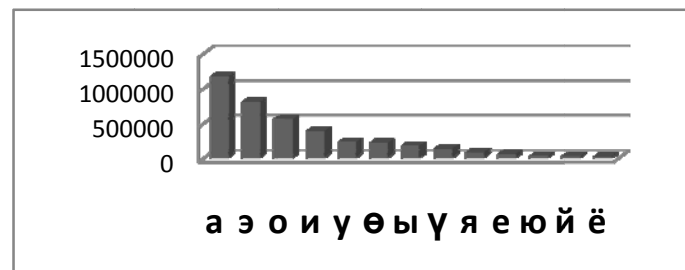
Зураг 4. Монгол хэлний эгшиг болон гийгүүлэгчийн хэрэглээ

Монгол хэлний үсгийн болон үеийн давтамжид тулгуурлан компьютерийн гарын үсгийн байрлалын жишиг загварыг гаргасан нь

Энэ хэсэгт судалгааны үр дүнгийн хэрэглээг компьютерийн гарын байрлалыг оновчлох жишээн дээр авч үзье. Зураг 5, Зураг 6-д монгол хэлний үсгийн давтамжийн судалгааг үзүүлэв. Хавсралт 1, Хавсралт 2, Хавсралт 3 –т SRILM хэрэгслийг² ашиглан гаргасан дан, хосолсон, гурвалсан үсгийн давтамжийн судалгааг оруулав.



Зураг 5. Монгол хэлний гийгүүлэгч үсгийн давтамж



Зураг 6. Монгол хэлний эгшиг үсгийн давтамж

²) SRI Language model хэрэгслийг <http://www.speech.sri.com/projects/srilm/> хаягаас татаж авч болно.

Одоо бидний хэрэглэж байгаа компьютерийн гарын үсгийн байрлалын эх сурвалжийг орос бичгийн машинаас улбаатай монгол(кирилл) бичгийн машины байрлалаас авсан байна. Тэгвэл энэ нь монгол хэлэнд, монгол хүнд таарсан, оновчтой байрлал биш болж таарна. Тиймээс монгол хэлний судалгаанд үсэг, үе, үгийн давтамжийг судалснаар үнэн бодитой гарын үсгийн байрлалыг шинэчлэн тогтоох боломж гарч ирж байгаа юм.

Бид Дворакийн хялбаршуулсан компьютерийн гарын үсгийн байрлалыг (Dvorak simplified keyboard)³ зохиосон зарчимд тулгуурлан загвар гаргалаа.

Дворакийн хялбаршуулсан үсгийн байрлалыг (Зураг 7) бидний хэрэглэж байгаа qwerty (бичгийн машины даруул дээрх үсгийн жишиг)-н үсгийн байрлалыг сольж болох механик хүндрэлийг тооцож зохиосон байна. Qwerty байрлалын сул тал нь: үсгийн нийлэмж нь хурууны хөдөлгөөнд тохиромжгүй, зөвхөн нэг гараар бичихийг шаарддаг, ихэнх бичиглэл нь хүмүүс голчлог баруун гараар бичдэг байхад зүүн гараараа бичихийг шаарддаг, нийт үсгийн 16% нь доод мөрд, 52% нь дээд мөрд, хамгийн гол мөр болох голын мөрд 32% л байдаг зэрэг юм.

Харин Дворак үсгийн давтамж болон хүний гарын физиологийн онцлогийг судалж, мөн qwerty-ийн байрлалд байгаа дутагдлуудыг хөнгөвчилсөн юм. Дворакийн байрлалын давуу талыг дурдвал: Үсгүүд нь 2 гарын ээлжлэн солигдох байдлаар бичихэд тохирсон (энэ нь айзам хэмнэлтэй) хурдыг нэмдэг, алдааг багасгаж, зүдэргээг багасгадаг. Дворакийн байрлалд эгшгүүд нь бараг бүгд дунд мөрийн зүүн талд хамгийн их хэрэглэгддэг гийгүүлэгчүүд нь баруун талдаа байрладаг.

Ингэж түгээмэл хэрэглэдэг үсэг болоод үсгийн дүрэм зэргийг харгалзан бичихэд хамгийн амархан, хурдан, үр дүнтэй байдаг. Энэ нь хаана сул хуруу байна түүгээр бичих эсвэл хүчтэй, чухал үүрэгтэй долоовор хуруу зэргээр бичих үсгүүдийг дунд мөрд байрлуулж, түүгээр голлон бичнэ гэсэн үг юм (Дворакийн байрлалд нийт бичсэн үсгийн 70% нь гол мөрд байгаа бол 22% нь дээд мөрд, 8% нь доод мөрд байрладаг).

Хамгийн бага хэрэглэгддэг үсгүүд нь хүрэхэд хол газар болох доод мөрд байдаг. Баруун гараараа илүү бичихээр зохиомжлогдсон. Учир нь ихэнх хүмүүс баруунаараа бичдэг.

~	!	@	#	\$	%	^	&	*	()	{	}	←
1	2	3	4	5	6	7	8	9	0	[]	Backspace	
Tab	"	<	>	P	Y	F	G	C	R	L	?	+	
Caps Lock	A	O	E	U	I	D	H	T	N	S	-	Enter	
Shift	:	Q	J	K	X	B	M	W	V	Z	Shift		
Ctrl	Win Key	Alt							Alt Gr	Win Key	Menu	Ctrl	

Зураг 7. Дворакийн хялбаршуулсан компьютерийн гарын үсгийн байрлал

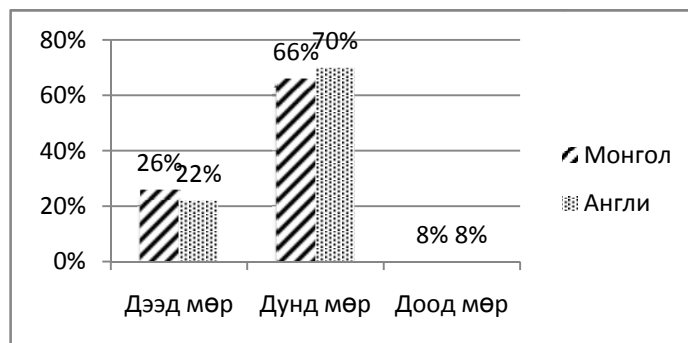
Дээрх зарчмыг баримтлан монгол хэлний дан, хосолмол, гурвалсан үсгийн давтамжид тулгуурлан монгол гарын байрлалыг тогтоолоо. (Зураг 8-г үзнэ үү)

³) Дворакийн байрлал нь Англи хэлэнд зориулан 1932 онд бүрэн зохиогдож дуусаад 1936 оюуны өмчийн эрхийг авсан.



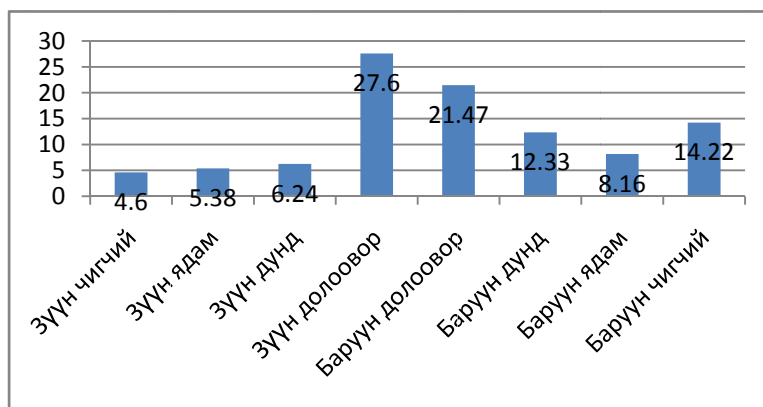
Зураг 8. монгол гарын байрлал

Энэхүү байрлалыг Дворакийн англи байрлалтай харьцуулахад ерөнхийдөө ойролцоо хувь гарсан байна. Энэ нь нийт үсгийн 66-70 хувийг компьютерийн гарын дунд мөрд, 22-26 хувийг дээд мөрд, 8 хувийг доод мөрд байрлуулсан гэсэн санаа юм.



Зураг 9. Дворакийн зарчимд тулгуурласан гарын дээд, дунд, доод мөр дэх монгол болон англи үсгийн байрлалын харьцуулал

Харин Зураг 10-т компьютерийн гарын байрлалын монгол үсгийн баруун болон зүүн гарын хуруу бүрт ногдож байгаа ачааллыг товчлуурт тавигдсан үсэгт нь үндэслэн үзүүлээ. Зургаас харахад баруун болон зүүн гарын долоовор хуруу бичихэд хамгийн их үүрэгтэй байгаа нь харагдаж байна. Нийтэд нь үзэхэд баруун гарын хуруунууд нийт ачааллын 56 хувийг, зүүн гарын хуруунууд 44 хувийг бичихэд оролцож байна. Энэ нь ихэнх хүн баруун гараараа бичдэгтэй холбож болно.



Зураг 10. Үсгийн хуруу бүрд ногдож байгаа ачаалал (хувиар)

Дүгнэлт

Энэхүү өгүүллийн хүрээнд кирил монгол 5 сая гаран үгтэй судалгааны эх хэрэглэгдэхүүн дээр тулгуурлан үе, үсгийн давтамжийн судалгааг хийлээ. Үүний үр дүнд хамгийн их давтамжтай үе, үеийн загвар, задгай ба битүү үеийн давтамж, гийгүүлэгч ба эгшиг үсгийн давтамжийг гаргасан бөгөөд энэхүү судалгаан дээр тулгуурлан монгол хэлний компьютерийн гарын жишиг загварыг зохиолоо.

НОМЗҮЙ

1. Амоголов Д., Современный бурятский язык, Улан-Удэ, 1958.
2. Базаррагчаа М., Утгын чуулган, Уб, 2002.
3. Баянсан Ж. Төмөртоого Д., Хэлшинжлэлийн нэр томъёоны зүйлчилсэн тайлбар толь, Уб, 1995.
4. Бураев И.Д., Становление звукого строя бурятского языка, Новосибирск, 1987, 184с.
5. Владимирцов Б.Я., Сравнительная грамматика монгольского письменного языка и халхаского наречия, Л, 1929, 437.
6. Дамдинсүрэн Ц., Осор Б. Монгол үсгийн дүрмийн толь, Уб, 1983.
7. Лувсанвандан Ш., ОЦМХ-ний бүтэц, Бүтээлийн чуулган V боть, Уб, 2010.
8. Лувсанвандан Ш., ОЦМХ-ний бүтэц, Авиа авиалбар хоёр нь Бүтээлийн чуулган V боть, Уб, 2010.
9. Ринчен Б., Монгол бичгийн хэлний зүй, II дэвтэр, 1966, 79-82 .
10. Санжаа Ж, Надмид Д., Орчин цагийн монгол хэлний авиазүй, Уб, 1993, 180-183.
11. Санжаа Ж, Надмид Д., Монгол хэлний авиазүй ба авиалбарзүй, Уб, 2008, 229-231.
12. Сайт: http://en.wikipedia.org/wiki/Dvorak_Simplified_Keyboard
13. Төмөртоого Д., Монгол хэлний үеийн бүтэц, хувьсал, SM.13(21): 1987, 83-92.
14. Төмөртоого Д., Монгол хэлний түүхэн хэлзүй-I, Уб, 1992, 173-179.
15. Энхжаргал, М. Пүрэв Ж, Нанзадрагчаа Д., Электрон бичвэрт суурилсан монгол хэлний үгийн болон үсгийн давтамжийн судалгаа, Хүрэлттоогоот-2012, Уб, 2012, 52-58 тал.
16. Graeme Kennedy., *An Introduction to corpus linguistics*, 1998.
17. Hattori Sh., *Phonemic Structure Of Mongol*, 1951, 68-102.

RESUME

In the scope of this study , we focused on the Mongolian letters and syllables frequency analysis. There are 11,123,714 syllables found in Corpus of Mongolian language which has more than 5 million words. Thus letter and syllable pattern frequencies are counted.

In conclusion, we designed Mongolian simplified keyboard layout based on the frequency analysis of the Ngrams (one to three).

ХАВСРАЛТ 1.

Монгол хэлний үсгийн давтамж(Unigram)

C1	Freq
а	3751001
н	2537142
э	2360169
г	2004018
л	1909750
р	1639747
х	1497453
о	1475742
д	1399588
и	1387090
й	1331955
у	1171869
т	1063784
ү	995238
с	925056
ө	847392
б	788327
м	694732
ж	536974
в	402827
ы	386175
ч	378931
з	321199
ш	259647
ь	223668
ц	218504
я	163813
е	116350
ю	74441
к	57684
ё	41719
п	33511
ф	8317
щ	100

ХАВСРАЛТ 2.

Монгол хэлний хосолмол үсгийн давтамж (Bigram)

#	C1	C2	Freq
1	и	й	586962
2	г	а	422601
3	э	э	391690
4	й	н	378207
5	б	а	352480
6	у	у	319343
7	э	р	317807
8	о	л	298763
9	г	э	291648
10	э	н	290140
11	л	а	267654
12	д	а	263350
13	х	а	255050
14	?	?	243218
15	с	а	236878
16	у	л	230665
17	о	н	223508
18	т	а	221467
19	э	г	219154
20	й	г	210998
21	н	а	208544
22	э	л	208009
23	ы	н	206893
24	д	э	201482
25	н	э	200833
26	х	э	196553
27	л	э	179894
28	б	о	177905
29	о	р	177736
30	т	э	177545
31	и	л	167486
32	?	й	156617
33	э	д	154457
34	?	р	153971
35	о	о	151460

Хавсралт 3.

Монгол хэлний гурвалсан үсгийн давтамж (Trigram)

#	C1	C2	C3	Freq
1	а	а	р	93967
2	а	й	г	85603
3	а	г	а	69497
4	а	й	н	67394
5	а	р	а	60512
6	а	а	н	60221
7	а	н	а	57407
8	а	р	и	54352
9	а	р	г	51862
10	а	а	с	49637
11	а	л	т	48491
12	а	ж	и	40733
13	а	л	а	38760
14	а	а	д	37103
15	а	й	х	37046
16	а	н	д	35256
17	а	с	а	34743
18	а	а	г	33482
19	а	л	д	32900
20	а	н	г	30972
21	а	й	д	30816
22	а	т	а	29400
23	а	д	а	27158
24	а	р	д	27065
25	а	в	а	26516

26	а	а	л	25749
27	а	р	ы	25558
28	а	н	ы	25413
29	а	й	с	24438
30	а	х	а	23971
31	а	л	г	23835
32	а	г	д	23668
33	а	й	л	22482
34	а	х	и	20241
35	а	ш	и	18726
36	а	л	б	18705
37	а	м	ж	17117
38	а	й	р	16724
39	а	г	ч	16471
40	а	р	т	16172
41	а	н	х	15732
42	а	м	ь	15602
43	а	р	ч	15576
44	а	м	а	15572
45	а	з	а	15429
46	а	л	ы	15372
47	а	а	т	14847
48	а	х	у	14482
49	а	м	г	14016
50	а	н	и	13285

Е-мэйл: altangerel@num.edu.mn,
enkhjargal.m@gmail.com