

Монгол хэлний үг, үе, үсгийн давтамжийн судалгаанд суурилан компьютерийн гарын жишиг байрлалыг зохиосон нь

Ч.Алтангэрэл*, М.Энхжаргал**

*Монгол улсын их сургууль, Мэдээллийн технологийн сургууль,
Компьютер, мэдээллийн технологийн тэнхим
golden_11@yahoo.com,

**Монгол улсын боловсролын их сургууль, Багшийн сургууль,
Хүмүүнлэг, нийгмийн ухааны тэнхим
enkhjargal.m@gmail.com

Хураангуй—Хэлийг цахимаар боловсруулахад зайлшгүй хэрэглэгдэх чухал зүйл бол тухайн хэлний мэдлэгийг шингээсэн үнэн бодит, их хэмжээний эхийн сан буюу цахим сан (corpus) юм. Цахим санд тулгуурлан хэл шинжлэлийн олон сонирхолтой судалгаа, дүгнэлтүүдийг хийж болох бөгөөд эдгээр судалгааны нэг нь тухайн хэлэн дэх үг, үе, үсгийн давтамжийн судалгаа юм. Үеийн давтамжийн судалгааг цахимаар хийх нь ихээхэн цаг, хүч хөдөлмөр хэмнэсэн ажил бөгөөд богино хугацаанд, үнэн зөв, цэгцтэй мэдээлэл, үр дүнг гаргадгаараа онцлог юм.

Уг өгүүлэл нь монгол хэлний цахим санд тулгуурлан монгол хэлний үг, үе, үсгийн давтамжийг судалсан талаар болон судалгаанаас гарсан үр дүн, түүнд тулгуурлан компьютерийн гарын байрлалын жишиг байрлалыг зохиосон тухай өгүүлнэ.

Keywords—*Корпус хэл шинжлэл; цахим хэл шинжлэл; монгол хэлний үе;*

I. УДИРТГАЛ

Корпус хэл шинжлэлийн ач холбогдол нь их бөгөөд олон талын судалгаанд хэрэгтэй. Ялангуяа хэл шинжлэлд цагаан толгойн үсгийн хэрэглэгдэх давтамж; үгсийн аймгийн өгүүлбэр дэх үүрэг, хэрэглээ; дагаврын чадамж, тархац; оносон хайлгаар 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. Корпус суурилсан хэл шинжлэлийн судалгааг дотор нь уран зохиолын, үгийн сангийн, нутгийн аялгууны, боловсролын, хэлзүйн гэж ангилна[Graeme Kennedy, 1998,13-17].

Тухайлбал, уран зохиолын судалгаанд Библийн үг хэллэгийг судалсан судалгаа, тодорхой нэг зохиогчийн (Жишээ нь Шекспир) уран бүтээлийн үг хэллэгийн онцлогийг судлах гэх мэт зорилгоор хийдэг байна. 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 хэллэг (collocations) хамгийн их хамт хэрэглэгдсэн байгааг тогтоожээ.

Мөн корпус суурилсан судалгааны гол зорилтуудын нэг нь сурган хүмүүжүүлэх арга ажиллагаанд хэрэглэх туслах материалыг бий болгох, тухайлбал, хоёрдогч хэлийг сурч байгаа хүнд зориулан тухайн хэлэнд хамгийн их хэрэглэгддэг үгийн давтамжийн жагсаалтыг гаргах зэрэг юм[М.Энхжаргал, Ж.Пүрэв, Д.Нанзадрагчаа, 2012].

Дээр дурдсанчлан хэлний судалгаанд тодорхой хэрэгцээ шаардлагын улмаас корпус суурилсан статистик судалгаа чухал байр суурь эзэлдэг билээ. Монгол хэлний 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 холбогдолтойгоос [Д.Төмөртогоо, 1986] гадна бидэнд монгол хэлний үеийн судалгааг хийх дараах шаардлага тулгарсан юм. Үүнийг дурдвал:

A. Нэгдүгээр шаардлага

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

B. Хоёрдугаар шаардлага

Үсэг хоорондын хамт хэрэглэгдэх давтамжийг судалснаар компьютерийн гарын үсгийн байрлалыг зөв тогтооход чухал. Тухайлбал, *бай-, ай-* гэсэн үг, үе монгол хэлэнд их давтамжтай байвал компьютерийн гарын үсгийн байрлалд ойрхон байрлуулах жишээтэй. Иймэрхүү тохиолдол орос үсгийн гарын байрлалд байдаг. Жишээ нь: *про* гэсэн үсгийн хослол орос хэлэнд их тохиолддогийг харгалзан уг 3 үсгийн хослолыг компьютерийн гарын үсэгт зэрэгцүүлэн байрлуулсан байна.

II. ЦАХИМ САНД ТУЛГУУРЛАСАН МОНГОЛ ХЭЛНИЙ ҮЕИЙН СУДАЛГАА

Авиан зүйн үүднээс монгол хэлний үеийн бүтцийн тухай судалгаанд акад. Б.Я.Владимирцов CV, SVC, VC гэсэн 3 янзын бүтэцтэй үе [Владимирцов Б.Я., 1929]., проф Ш.Хаттори CV [Hattori Sh, 1951], мөн оросын эрдэмтэн Д.Амоголов, И.Д.Бураев, акад Ш.Лувсанвандан нар V, VC, CV, SVC гэсэн 4 янзын бүтэцтэй үе байна 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 [Төмөртогоо Д, 1992].

Үеийн хувьслыг түүхэн үүднээс авч үзсэн судалгааг акад Д.Төмөртогоо [Д.Төмөртогоо, 1987,1992], проф М.Базаррагчаа [М.Базаррагчаа, 2002] нар хийж монгол хэлний үеийн тухай судалгааг улам баяжуулсаар байна.

Амьсгалын нэг түрэлтээр хэлэгдэх нэг буюу хэдэн авиаг үе гэх бөгөөд ман-дуул-сан, 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].

“Эртний монгол хэлэнд Э, ГЭ, ЭГ, ГЭГ гэсэн үеийн 4 хэлбэр байх ба ГЭГТ гэсэн үе хааяа тохиолдоно[Санжаа Ж., Надмид Д, 2008]” гэжээ.

Харин хэлний бодит материал дээр тулгуурлан монгол хэлний кирилл 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” [Санжаа Ж., Надмид Д, 2008] гэж үзсэнийг эс тооцвол бараг байхгүй байна.

Энэхүү судалгааг монгол хэлний бодит материал буюу цахим бичвэрт суурилан нарийн тооцон гаргасан юм. Үүнийг МУИС-ийн МТС-ийн Компьютер хэл шинжлэлийн судалгааны төвд төвлөрөн хийсэн Монгол хэлний 5 сая орчим үгээс бүтэх материалын санд тулгуурлан гаргасан болно. Энэхүү цахим сан нь давхардаагүй тоогоор 170000 орчим үг байна. 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.

Монгол хэлний үеийн давтамжийг гаргахдаа кирилл үгийг үелэх зарчмын дагуу үгийг зааглан хуваагаад эгшиг, гийгүүлэгчээр нь солин автоматаар гаргасан болно. Уг судалгаанд хэрэглэсэн энэхүү 5 сая гаран үлтэй кирилл монгол хэлний эхийн санг хойшид *судалгааны эх* гэдгээр ойлговол зохино. Судалгааны эхээс гарсан давтамжийн үр

дүнг монгол хэлний бодит онцлогоор төлөөлүүлэн үзэж болно.

Ингээд тооцооллын дүнд судалгааны эхэд 18,088 төрлийн нийт 11,123,714 үе байна. Үүнээс үзэхэд монгол хэлний нэг үг дунджаар 2 үеэс бүтдэг болохыг харж болно. Үеийн давтамжийг гаргахдаа AntConc¹ програмыг ашигласан.

Хүснэгт I. Хамгийн их давтамжтай үеийн жагсаалт

#	Давтамж	Үе	Жишээ нь
1	178375	бай	өр нүүлж *бай гаа *улс *гүм ний *
2	143260	а	а жил * *аж *а хуйн *бай гуул лал тыг
3	140774	га	ар га *хэм жээ *авч *бай вал
4	108008	сан	*өм нө *тул гар сан *зо рил тыг
5	105830	на	юм *шиг *са нар ла на
6	105509	нь	харь яал лаар *нь *ху ваа ри лах
7	102691	нэ	ши нэ *ша ганд *гар гах *шаард ла га
8	101096	гүй	*ча дах гүй *бай на *
9	97765	о	*о лон *ний гийн *бай гуул ла га
10	91318	ны	ан хан *шат ны *бай гуул ла гуу дын



Зур.1. Хамгийн их давтамжтай үе

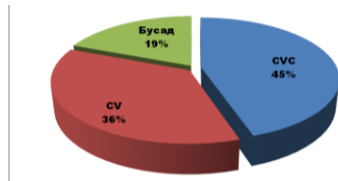
Энэхүү судалгааны эхэд дурдсанчлан 7 төрлийн загвартай үе тохиолдож байна. Энэ нь кирилл бичгийн үгийг үелсэн үеийн хэлбэрийн судалгаа юм. Уг үеийн загварыг гаргахдаа гадаад үг, цэг тэмдэг, дэс дугаар, товчилсон үг, бүтээвэр зэргийг (3500 орчим) оруулаагүй болно.

Хүснэгт II. ЦАХИМ САНД ТОХИОЛДОЖ БУЙ ҮЕИЙН ЗАГВАР

#	Давтамж	Үеийн загвар	Жишээ
1	5,006,222	SVC	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

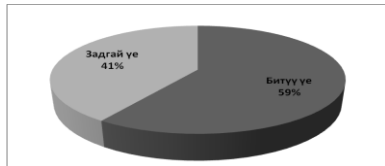
Мөн судалгаанаас харахад SVC гэсэн үе эхэд хамгийн их хувийг эзэлж байгаа нь харагдаж байна. Энэ нь “Санжаа Ж, Надмид Д (2008) нарын SVC гэсэн үе монгол хэлэнд зонхилох суурь буюу 50 гаруй хувийг эзэлж байна” гэсэн судалгаатай таарч байгаа юм.

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



Зур.2. Үеийн төрлийн эзлэх хувь

Цаашилбал, битүү үе буюу гийгүүлэгчээр төгссөн үе задгай үеэсээ их тохиолдож байгаа нь гийгүүлэгч үсгийн хэрэглээ эгшгээс их байгааг харуулж байна.



Зур.3. Битүү болон задгай үеийн эзлэх хувь

III. ЦАХИМ САНД ТУЛГУУРЛАСАН МОНГОЛ ХЭЛНИЙ ҮГИЙН ДАВТАМЖИЙН СУДАЛГАА

Таван сая орчим үгтэй монгол хэлний цахим санд тулгуурлан үгийн давтамжийг судалж монгол хэлэнд хамгийн их давтамжтай хэрэглэгдэж байгаа 200 үгийн жагсаалтыг гаргалаа.

Хүснэгт III. Монгол хэлний хамгийн их давтамжтай 200 үг

#	Давтамж	Үг
1	97361	бай-
2	72002	гэ-
3	62634	нь
4	33723	бол-
5	25294	юм
6	23950	ч
7	20060	би
8	17834	л
9	15780	нэг/нэгэн
10	15489	яв-

Эдгээр 200 үгийг Хүснэгт 1 харж болно. Уг судалгаа нь монгол хэл заах аргад чухал ач холбогдолтой бөгөөд ялангуяа монгол хэл эхлэн сурч байгаа гадаадын хүнд монгол хэлэнд хамгийн их хэрэглэгддэг эдгээр 200 үгийг нүдлэн таниулж, тайлбарласнаар эдгээр үгсийг ашиглан ойр зуурын хэрэглээнд харилцан ярилцах, богино хэмжээний эх зохиолдох гэх мэтээр зүгшрүүлэхэд ашиглаж болно.

Дараа нь уг гаргасан үгсээ Brown corpus-т хамгийн их хэрэглэгдэж байгаа 200 үгтэй харьцуулав.

Brown corpus нь америкийн англи хэлийг судлах зорилгоор 1960-аад онд зохиогдсон бөгөөд 1 сая орчим үгээс бүтнэ. Уг харьцуулалтын үр дүнд монгол, англи хоёр хэлний тус бүр 200 үгийн 91 нь буюу тавь орчим хувь нь утга агуулгаараа хоорондоо таарч байна. Дээрх хоёр корпусаас их давтамжтай үгийн жагсаалт гаргахдаа *AntConc 3.2.1* програмыг ашигласан болно.

Хүснэгт IV. Монгол болон англи бичвэрийн санд орсон хамгийн их давтамжтай үгсийг харьцуулбал

Монгол хэлэнд буй	Давхардаж буй	Англи хэлэнд буй
Монгол мал, ахуй, нам, төр, МАХН, УИХ, сэтгэл, хот, сум, аймаг, сар, эрх, ам, нэр, мянга, даа, шүү	бай-/be, нэг/one, яв-/went, олон/many, ус./water, ямар/what, гар/hand, буюу/an, d, би/I, бид/we, гэр/home.	American, a, an, the, Mr, Mrs, school, world, use, against, life

Дээрх Хүснэгт 4-өөс харахад давхардахгүй байгаа үгс нь хоёр орны соёл, ахуй, уламжлалаас үүдэлтэй байгаа бөгөөд энэ нь монгол хэлэнд мал, ахуй, аймаг, сум, хот гэх мэт үгээр илэрч байхад, англи хэлэнд Mr, Mrs гэх мэт үгээр илэрч байна. Мөн монгол хэлэнд улс төр, намтай холбоотой үгс элбэг тохиолдож байгааг нам, төр, МАХН, УИХ гэх мэт үг гэрчилнэ. Харин хоёр хэлний корпус давтамжаараа давхардаж байгаа үгсэд, нэг(one), хоёр(two), гурав(three) гээд нэгээс гурав хүртэлх тооны нэр; би(I), чи(you), бид(we), тэд(they), хэн(who), ямар, юу(what), яаж (how) хаана(where) гэх мэт биеийн болон асуух төлөөний үг; гэвч(but), учир (because), мөн, ба, буюу, болон(and), тэгээд(so) зэрэг холбоос; ёстой(must), магадгүй(may) зэрэг баймж утгат үгс зэрэг багтаж байна.

Тэргүүн өгүүлэл нь хамгийн чухал сэдэвтэй, тухайн цаг үетэй холбоотой үйл явдлыг мэдээлсэн байдаг бөгөөд уриалан дуудсан, хүмүүжүүлсэн, зааж сургасан өнгө аястай, ойролцоогоор 1000 орчим үгээс бүтэж Үнэн сонины хамгийн эхний нүүрэнд тогтмол хэвлэгддэг байв. Үнэн сонины тэргүүн өгүүлэлд давтагдаж буй үгс нь 1984-1989 оны улс орны уур амьсгал, улс төртэй холбоотой үгс байгааг аж ахуй, ажил, улс, нам, их, олон гэх мэт үгсээс харж болохоор байна.

Харин хуулийн эхэд Монгол улсын үндсэн хууль болон бусад хууль нийлээд 144 төрлийн хууль орсон бөгөөд дундаж үгийн тоо 4000 орчим байна. Хуулийн эхэд давтагдаж байгаа үгс нь дугаар, зүйл, хууль, сар, улс, он, эрх гэх мэт албан бичиг, хуулийн нэр томъёо мөн болно.

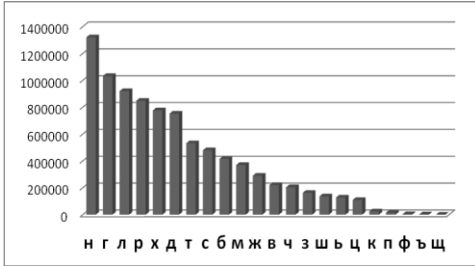
Уран зохиолын эхэд монгол орны ахуй, уламжлалтай холбоотой үг хэллэг түгээмэл орсноос гадна зохиолын баатруудын нэрний давтамж их байгаа нь ажиглагдлаа. Тухайлбал, зохиолч Ч.Лодойдамбын “Тунгалаг Тамир” романы гол баатар болох Эрдэнэ 1285 удаа, Итгэлт 860 удаа, Д.Намдагийн “Цаг төрийн үймээн” романы Данзан 725, Балданцэрэн 590 удаа давтагдан орсон байна.

Сонин нийтлэлийн найруулгын эхэд нийт 2164676 үг байгаагаас 120695 нь гадаад үг байгаа нь сонин нийтлэлийн эхийн 6 хувийг эзэлж байна. Тухайлбал, *CD, Pentium, PR, school, SHOP, TOEFL, TV, UB* гэх мэт.

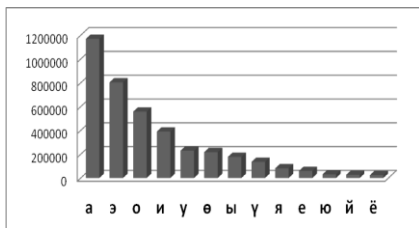
IV. ЦАХИМ САНД ТУЛГУУРЛАСАН МОНГОЛ ХЭЛНИЙ ҮСГИЙН ДАВТАМЖИЙГ СУДАЛГАА

Судалгааны эхэд давхардаагүй тоогоор 170 000 орчим үг байгаа бөгөөд үүнд уран зохиолын эхээс роман, тууж, өгүүлэл; сонин нийтлэлийн эхээс мэдээ, тэргүүн өгүүлэл (Үнэн сонины), нийтлэл, ярилцлага, сурвалжлага,

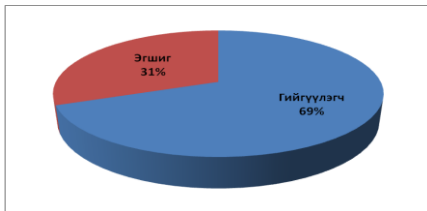
өгүүлэл; хуулийн эхээс үндсэн хууль болон бусад хуулийг сонгон авсан болно. Хугацааны хувьд эдгээр бичвэр нь 1920-2008 онд хамаарч байгаа бөгөөд гадаадын болон эртний уран зохиолоос бичвэр ороогүй болно. Ингээд үсгийн давтамжийг гаргахад *н* үсэг хамгийн монгол хэлэнд хамгийн их давтамжтай байна. Монгол хэлний үсгийн давтамжийг Зураг 4, 5, 6-д үзүүлээ.



Зур.4. Монгол хэлний гийгүүлэгч үсгийн давтамж



Зур.5. Монгол хэлний эгшиг үсгийн давтамж



Зур.6. Эгшиг болон гийгүүлэгч үсгийн тархац

V. МОНГОЛ ХЭЛНИЙ ҮГ, ҮЕ, ҮСГИЙН ДАВТАМЖИД СУУРИЛАН КОМПЬЮТЕРИЙН ГАРЫН ЖИШИГ БАЙРЛАЛЫГ ЗОХИОСОН НЬ

Дээрх монгол хэлний үг, үе, үсгийн давтамжийн судалгааны үр дүнгээр компьютерийн гарын байрлалыг жишиг загвар зохиолоо. Үүнийг SRILM хэрэгслийг² ашиглан гаргасан.

Одоо бидний хэрэглэж байгаа компьютерийн гарын үсгийн байрлалын эх сурвалжийг авч үзвэл орос бичгийн машинаас улбаатай монгол(кирилл) бичгийн машины байрлалаас авсан байна. Үнэхээр орос хэлэнд зохицсон орос үсгийн байрлалаас авсан гэж үзвэл одоо бидний хэрэглэж байгаа гарын үсгийн байрлал нь монгол хэлэнд, монгол хүнд таарсан, оновчтой байрлал биш болж таарна. Тиймээс монгол хэлний судалгаанд үсэг, үе, үгийн давтамжийг судалснаар үнэн бодитой гарын үсгийн

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

байрлалыг шинэчлэн тогтоох шаардлага гарч ирж байгаа юм.

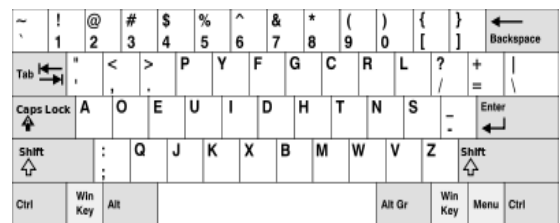
Бид энэхүү жишиг загварыг тогтоохдоо Дворакийн хялбаршуулсан компьютерийн гарын үсгийн байрлалыг (Dvorak simplified keyboard)³ зохиосон зарчимд тулгуурласан.

Дворакийн хялбаршуулсан үсгийн байрлалыг (Зураг 7) одоо бидний хэрэглэж байгаа qwerty (бичгийн машины даруул дээрх үсгийн стандарт)-н үсгийн байрлалыг солиод хэрэглэж болохоор механик хүндрэлийг тооцож зохиосон байна. Qwerty байрлал нь үсгийн нийлцүүд нь хурууны хөдөлгөөнд тохиромжгүй, зөвхөн нэг гараараа бичихийг шаарддаг, ихэнх бичиглэл нь хүмүүсийн хувьд сул гар болох зүүн гараараа бичихийг шаарддаг, нийт бичсэн үсгийн 16% нь доод мөрөнд, 52% нь дээд мөрөнд, хамгийн гол мөр болох голын мөрөнд 32% л байдаг зэрэг сул талтай юм.

Гэтэл Дворак нь 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 зүүн голын мөрөнд байдаг ба хамгийн их хэрэглэгддэг гийгүүлэгчүүд нь баруун талдаа байрладаг.

Хамгийн их хэрэглэгддэг үсэг болоод үсгийн дүрэм зэргийг харгалзан бичихэд хамгийн амархан, хурдан, үр ашиг ихтэй байдаг. Энэ нь хаана сул хуруу байна түүгээр бичих эсвэл хүчтэй, гол үүрэгтэй долоовор мэт хуруугаар бичих үсгүүдийг голын мөрөнд байрлуулах, тэдгээрээр голлон бичнэ гэсэн үг юм(Дворакийн байрлалд нийт бичсэн үсгийн 70% нь гол мөрөнд байгаа бол 22% нь дээд мөрөнд, 8% нь доод талын мөрөнд байрладаг байна).

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



Зур.7. Дворакийн хялбаршуулсан компьютерийн гарын үсгийн байрлал

Үүнээс үүдээд хоёр гарын ээлжлэн солигдох байдлаар бичихэд тохируулан баруун талд гийгүүлэгч үсэг, зүүн талд эгшиг үсгийг байрлуулах, хамгийн их хэрэглэгддэг үсгийг хамгийн гол үүрэгтэй хуруугаар (долоовор) бичих, хамгийн их хэрэглэгддэг үсгийг компьютерийн гарын голын мөрөнд, бага хэрэглэгддэг

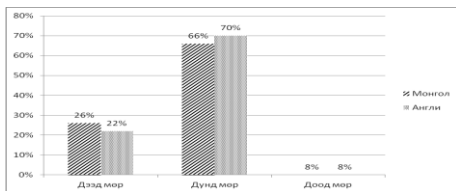
³ Дворакийн байрлал нь Англи хэлэнд зориулан 1932 онд бүрэн зохиогдож дуусаад 1936 оюуны өмчийн эрхийг авсан.

үсгийг захад байрлуулах зэрэг зарчмыг баримтлан монгол хэлний дан, хосолмол, гурвалсан үсгийн давтамжид тулгуурлан Монгол хэлний компьютерийн гарын үсгийн байрлалыг зохиолоо. Уг хялбаршуулсан байрлалын зургийг Зураг 8-д үзүүлээ.



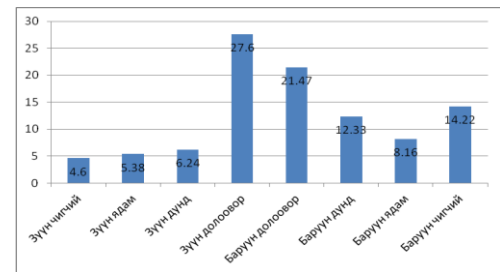
Зур.8. Дворакийн хялбаршуулсан гарын үсгийн байрлалыг зохиох зарчимд тулгуурлан зохиосон кирилл монгол үсгийн байрлал

Зураг 9 харахад Дворакийн зарчимд тулгуурласан компьютерийн гарын дээд, дунд, доод мөрөн дэх монгол болон англи үсгийн байрлалыг харьцуулахад ерөнхийдөө ойролцоо хувь гарсан байна. Энэ нь нийт бичигдэж байгаа үсгийн 66-70 хувийг компьютерийн гарын голын мөрөнд, 22-26 хувийг компьютерийн гарын дээд мөрөнд, 8 хувийг компьютерийн гарын доод мөрөнд байрлуулсан гэсэн санаа юм.



Зур.9. Дворакийн зарчимд тулгуурласан компьютерийн гарын дээд, дунд, доод мөрөн 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.



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

VI. ДҮГНЭЛТ

Энэхүү өгүүллийн хүрээнд кирилл монгол 5 сая гаран үгтэй судалгааны эх хэрэглэгдэхүүн дээр тулгуурлан үг, үе, үсгийн давтамжийн судалгааг хийлээ. Үүний үр дүнд хамгийн их давтамжтай үе, үг, үсгийн давтамжийг гаргасан бөгөөд энэхүү судалгаан дээр тулгуурлан монгол хэлний компьютерийн гарын жишиг загварыг зохиолоо.

Цаашид бидний зохиосон энэхүү гарын байрлалыг нийтэд хэрэглэж заншсан одоогийн гарын байрлалтай харьцуулан, хэр үр ашигтай байгаа талаар судалгаа хийх шаардлагатай.

ТАЛАРХАЛ

Энэхүү ажлыг хийхэд тусалж дэмжин, монгол улсад компьютер хэл шинжлэлийн судалгааг хөгжүүлэх үйлсэд бодитой хувь нэмэр оруулж яваа МУИС-ийн МТС-ийн КХШСТ-ийн эрхлэгч проф Ж.Пүрэв багшдаа чин сэтгэлийн талархал илэрхийлье.

НОМЗҮЙ

- [1] John Sinclair, *Corpus, Concordance, Collocation*, Oxford New York, 1996.
- [2] Graeme Kennedy, *An introduction to corpus linguistics*, London and New York, 1998.
- [3] Chagnaa Altangerel and Jaimai Purev: *Web as a corpus*, Mongolian Studies Research papers, vol. XXXI (328), 2010, ISSN 1997-1826 (in Mongolian)
- [4] Пүрэв Ж. нар, *Монгол хэлний тулгуур цахим байгууламж*, МУИС, Азийн судалгааны төв, Улаанбаатар, 2006.
- [5] Эрдэнэбаатар А., Пүрэв Ж нар *Персональ компьютерт монгол хэлний цагаан толгойн үсэг байрлуулах нь*, Техник технологийн мэдээ, IV, 1989 он.
- [6] Enkhjargal M, Purev J, *Mongolian orthography analysis based on Electron text*, Mongolian studies-New Approaches International Conference, Proceedings, p101-102, 2012.