

МАШИНЫ ОРЧУУЛГЫН ХӨГЖИЛ БА ХЭРЭГЛЭЭ

Ж.Эрдэнэбат¹

Товч утга: Энэхүү өгүүлэлд орчин үед хурдацтай хөгжиж буй машины орчуулга ба компьютерт суурилсан орчуулга, түүний эрэлт хэрэгцээ, машины орчуулгын ажиллах горим, хэрэглээний хэлбэрүүд, хөгжлийн чиг хандлага, хүний орчуулга ба машины орчуулгын онцлог, давуу болон сул тал, орчуулгын сургалтад анхаарах зарим нэг асуудлыг дэвшүүлсэн болно.

Түлхүүр үг: машины орчуулга, компьютерт суурилсан орчуулга, орчуулгын програм хангамж, орчуулгын төрөл зүйл, машины орчуулгын эрэлт хэрэгцээ

Машины орчуулга

Машины орчуулга ба компьютерт суурилсан орчуулга нь хүний орчуулах үйл явцыг хөнгөвчлөхөд чиглэсэн онлайн ба онлайн бус горимын компьютерийн програм хангамжийг ашиглан, орчуулгын ажлыг гүйцэтгэдэг шинэ тутам хөгжиж буй орчуулгын хэлбэр юм.

Хүний тусламжгүйгээр төгс чанартай орчуулга хийх автомат машины орчуулгын систем хараахан бий болоогүй байгаа хэдий ч *Google*, *SYSTRAN*, *Microsoft* зэрэг томоохон компаниуд энэ салбарт ихээхэн хөрөнгө оруулалт хийж, тэр хэмжээгээр машины орчуулга хөгжсөөр байна. Тухайлбал янз бүрийн түвшинд програмчилагдсан 100 гаруй хэлний сантай “*Google Translate*” програмыг өдөр бүр 200 сая гаруй хүн 100 тэрбум орчим үг хэллэгийг орчуулахад хэрэглэж байна¹.

Машины орчуулгаас монгол хэл хоцорсонгүй. Монгол хэлийг “*Google Translate*” програмд оруулах ажилд Монголын тал *Google* корпорацитай идэвхтэй хамтран ажиллаж, 2013 оноос эхлэн монгол хэлийг тус үйлчилгээнд нэвтрүүлсэн байна. Эдүгээ “*Google Translate*” үйлчилгээг ашиглан интернет олон хэлний орчинд үнэ төлбөргүйгээр эх хэл рүүгээ мөн эх хэлнээсээ орчуулга хийх боломж бидэнд хэдийнээ бүрджээ.

“*Google Translate*” нь “*machine learning*” буюу өөрөө өөрийгөө сайжруулах горимын орчуулгын систем юм. Тиймдээ ч их хэмжээний мэдээллийн сан шаарддаг. Програм нь 10 сая ба түүнээс дээш тооны хос өгүүлбэр /sentence pair/ орсон хэлний хувьд орчуулгын чанар тун сайн байгаа бөгөөд үүнд англи, испани, герман, франц зэрэг найман хэлний орчуулгыг дурдаж болно. “*Google Translate*”-ийн монгол хэлний орчуулга чанарын хувьд хангалтгүй байгаа нь мэдээллийн сан бага байгаатай холбоотой. Гэхдээ тогтмол хэрэглэж, алдааг залруулах тутам ‘*machine learning*’

¹ МУИС, ШУС, ХУС-ын ИБАСТ-ийн багш

систем тодорхой алгоритмын дагуу өөрөө суралцаж, мэдээллийн сангаа баяжуулах тул монгол хэлний орчуулга улам бүр төгөлдөржинө гэдэгт итгэлтэй байна.

2016 оны арваннэгдүгээр сард *Google* корпораци “*Google Translate*” програмынхаа орчуулгын чанарыг сайжруулах зорилгоор электрон “оюун ухаан–нерв” горимын системийг боловсруулж, GNMT буюу *Google*-ийн электрон нерв бүхий машины орчуулгыг эхний ээлжинд англи, хятад, франц, герман, япон, солонгос, португал, испани, турк зэрэг найман хэл дээр нэвтрүүлжээ. Эдгээр хэл нь “*Google Translate*” програм ашиглалтын 35% орчим хувийг эзэлж байгаа аж². Удалгүй 2017 оны гуравдугаарт сард дахин орос, хинди, вьетнам гурван хэл дээр нэвтрүүлсэн байна. Мөн энэхүү програм нь дээрх найман хэл дээр өгүүлбэрийн түвшинд орчуулга хийх хэмжээнд боловсронгуй болсноос гадна орчуулгын түүхийг хуримтлуулах болсноор орчуулгын чанар улам бүр боловсронгуй болж, хүний орчуулгаас дутахааргүй түвшинд хүрсэн талаар судлаачид бичиж байна². *Google*-ийн хувьд энэ бүхэн зөвхөн эхлэл ажээ. Тус компани шинэ тутам боловсруулсан програмаа одоогоор зөвхөн дээрх найман хэл дээр ашиглах боломжийг бүрдүүлээд байгаа ч цаашид судалгаагаа гүнзгийрүүлэн “*Google Translate*” програмын бусад хэлний орчуулгад нэвтрүүлэх зорилгоор судалгаа хийж байна.

Мөн 1968 онд байгуулагдсан, машины орчуулгын чиглэлээр дагнасан үйл ажиллагаа явуулдаг *SYSTRAN* компани дүрэмд суурилсан машины орчуулгын програмаа улам бүр төгөлдөржүүлж, 2010 онд дүрэм ба статистик өгөгдөл хосолсон машины орчуулгын програмыг нэвтрүүлж, 130 гаруй хэлээр орчуулгын үйлчилгээ үзүүлж байна. Эдүгээ тус компани дэлхий даяар олон хэл хоорондын харилцааг хөгжүүлэх зорилгоор цэвэр электрон нерв горимын машины орчуулгын (Pure Neural Machine Translation) програмыг мөн боловсруулаад байна.

МУИС цахим орчуулгын хөгжилтэй хөл нийлүүлэн алхаж, англи–монгол онлайн орчуулгын системийг туршилтын журмаар ажиллуулжээ. Уг орчуулгын програм хангамжийн ажиллах горим нь дүрэмд суурилсан машины орчуулгын програмын горимоор орчуулгын бэлтгэл үе, орчуулах үе, үр дүнг нэгтгэх үе гэсэн үндсэн гурван хэсгээс бүрдэл болдог байна⁶. Тодруулбал орчуулгын бэлтгэл үед програм хангамж өгөгдлийг энгийн бичвэрт хөрвүүлж, орчуулах үед хэлзүй, өгүүлбэрзүй, найруулгазүйн олон шатлалт үйлдэл хийдэг бол үр дүнг буцаан нэгтгэснээр орчуулга бэлэн болдог байна.

Машины орчуулгын хөгжилд нөлөөлж буй хүчин зүйл бол орчуулгын өөрчлөгдөн шинэчлэгдэж буй эрэлт хэрэгцээ юм. Жишээ нь хүний орчуулсан, уламжлалт, чанартай орчуулгын эрэлт хэрэгцээнээс гадна чанарын хувьд харьцангуй доогуур, тухайлбал интернет орчинд тодорхой эх материалын ерөнхий утгыг олж авахад илүүтэй анхаарсан хэрэглэгчид олширсоор байна. Үүнээс гадна харилцааны энгийн хэлбэрүүд тухайлбал утас, электрон захидал, *viber*, *wechat* зэрэг онлайн горимын харилцааны програм, олон нийтийн газарт өрнөдөг өдөр тутмын яриа гэх мэт хэрэглээний олон янзын хүчин зүйлсийн улмаас машины орчуулгын эрэлт хэрэгцээ нэмэгдэж, чанарын хувьд улам бүр боловсронгуй болсоор байна. Үүний

нэгэн жишээ бол 2017 оны нэгдүгээр сард танилцуулсан Японы “Logbar” компанийн “ili” хэмээх авсаархан орчуулгын төхөөрөмж юм. Уг төхөөрөмж нь 50,000 үг хэллэгийн сантай бөгөөд интернетгүй орчинд ердөө 0.2 секундийн хурдтайгаар англи, япон, хятад хэлнээ хүний хэлсэн мэдээллийг төвөггүйхэн орчуулдаг байна. Тэр ч бүү хэл хэлний мэдлэг сайтай орчуулагч хүн орчуулгад тохирох оноосон нэршлийг сонгохдоо тайлбар толиноос илүүтэйгээр орчуулгын програм хэрэглэх, эх ноорог орчуулгаа боловсруулахдаа компьютерт суурилсан орчуулгын програм ашиглаж, дараагаар нь хянан тохиолдуулах хэлбэрийн орчуулах явц нэмэгдэж байгаа нь машины орчуулга шинэ түвшинд гарахад ихээхэн түлхэц үзүүлж байна.

Өсөн нэмэгдэж буй машины орчуулгын хэрэглээний дээрх хэлбэрүүдийг Жон Хатчинс **а. утга задлах** (орчуулагчдын орчуулгын програм хангамжийг ашиглаж буй хэлбэр), **б. утга ижилсэх** (эх материалын ерөнхий утгыг тайлахад илүүтэй анхаарсан хэрэглээний хэлбэр), **в. утга сэлгэх** (өдөр тутмын хэрэглээний хэлбэр), **г. мэдээллийн хүртээмж** (интернет, мэдээллийн технологийн түвшин дэх олон хэлний мэдээллийн сан) гэсэн дөрвөн зүйл болгон ангилжээ³.

Түүхэн хөгжил

Автомат орчуулгын системийн түүхийг сөхөж үзвэл компьютерийг зохион бүтээсэн тэртээ 1940 оноос эхэлжээ. Эхэн үедээ толь бичгээс өөрцгүй байсан уг систем 1980-аад оноос шинэ түвшинд хүрсэн цахим хэл шинжлэл болон олон үйлдэлт програм хангамжуудын тусламжтайгаар орчуулгын эхэд үгзүй, өгүүлбэрзүй, цаашлаад утгазүйн нарийвчилсан задаргаа хийх боломжтой болсон байна. Улмаар ижил нэр (*homonym*), утгын оновчтой холбоог тодорхойлох програм хангамжууд нэмэгдэж машины орчуулгыг шинэ түвшинд хүргэжээ.

Машины орчуулгыг эхэн үедээ холбогдох тоног төхөөрөмжийн зардлын улмаас засгийн газар, олон улсын байгууллага, цэргийн анги нэгтгэл ашиглаж байв. Жишээ нь 1970-аад оны үед “Systran” орчуулгын програмыг Оросын цэргийн зориулалттай шинжлэх ухаан, техникийн баримт сэлтийг англи хэлнээ орчуулах зорилгоор АНУ-ын агаарын цэргийн хүчин ашиглаж байв. “Systran” орчуулгын програмын орчуулгын чанарыг зарим тохиолдолд 90%-ийн гүйцэтгэлтэй гарч байсан тухай мэдээлэл байдаг⁴. Харин Европт машины орчуулгыг хамгийн анх ашигласан байгууллага нь Европын Комисс бөгөөд англи хэлнээс франц хэлнээ орчуулга хийх “Systran” орчуулгын програмыг 1976 онд суурилуулжээ. Энэ үеэс “Systran” орчуулгын програмыг орчуулгын зорилгоор ашиглаад зогсохгүй удирдах албан тушаалтнууд гадаад хэлээр хэлэх үг, албан бичгийн эх бэлтгэлдээ ашигладаг болсон байна.

1990-ээд он хүртэл машины орчуулгад ихээхэн хөрөнгө оруулалт хийсэн авч үр дүн муутай байсан бөгөөд машинаар орчуулсан эхийг орчуулагч дахин хянах зэргээр орчуулгын зардал буурахгүй байв. Гэвч машины орчуулгыг пүүс компаниуд зах зээлд нийлүүлэх бараа бүтээгдэхүүнийхээ сурталчилгаа болон техникийн баримт бичгийн орчуулгад түгээмэл хэрэглэх болсноор орчуулгын програм шинэ түвшинд гарах нөхцөл дахин бүрджээ. Учир нь техникийн бараа бүтээгдэхүүнийг дагалдах

гарын авлага болон бусад баримт бичиг нь агуулгын хувьд хэт их, богино хугацаанд олон хэл дээр нэгэн зэрэг орчуулагдахаас гадна бүтээгдэхүүний дараагийн шинэ загвар гарахтай холбоотойгоор цаг хугацаа, хөрөнгө мөнгө хэмнэх шаардлага зүй ёсоор тулгарч байжээ. Эдгээр хүчин зүйлийн улмаас техникийн материалын орчуулга нь хүний хүчин чадлаас давсан асуудал болж, томоохон компаниуд машины орчуулгыг хөгжүүлэхэд үлэмжхэн хөрөнгө оруулалт хийх болсон байна. Машины орчуулгыг амжилттай ашиглаж буй олон жишээнээс дурдвал Шведийн Эриксон корпорац *Logos* орчуулгын програм, АНУ-ын Форд, Женерал Моторс, *Xerox* корпорациуд *Systran* орчуулгын програм хангамжийг тус тус ашиглаж байна. Мөн томоохон компаниуд орчуулгын програм хангамжийг ашиглахдаа бараа бүтээгдэхүүнийхээ онцлогт нийцүүлэн хэлний хэм хэмжээг тохируулах оролдлого хийдэг байна. Тухайлбал *Xerox* корпорацийн бараа бүтээгдэхүүний техникийн баримт бичиг боловсруулах багийнхан *Systran* систем ашиглан орчуулгын эх бэлтгэхдээ олон улсын хэмжээнд зөвшөөрөгдсөн англи хэл-найруулгаар боловсруулах үүрэгтэй ажээ⁵. Тодруулбал өгөгдсөн нэр томъёо болон өгүүлбэрийн бүтцийг ашиглан бичдэг байна. Ингэснээр машины орчуулгын систем хоёрдмол утгаас зайлсхийх боломж бүрдэх бөгөөд нэгэн зэрэг олон хэл дээр, богино хугацаанд орчуулгын хувьд чанартай, техникийн баримт бичиг боловсруулдаг байна. Энэ мэтчилэн “техникийн тусгай хэл, найруулгын” давуу талыг пүүс компаниуд хэдийнээ хүлээн зөвшөөрч, хэрэглээ нь нэмэгджээ.

Машины орчуулгын горим

Ерөнхийдөө машины орчуулгын програм нь нэг хэлний үгийг нөгөө хэлний үгээр орлуулах горимоор ажилладаг. Гэвч хэрэглээний үед зөвхөн үгээр зогсохгүй хэлц, цаашлаад өгүүлбэрийг хоёр хэлнээ хамгийн оновчтой түвшинд тохируулан оноох шаардлага зүй ёсоор урган гардаг тул дан үг оноолтоор чанартай орчуулга гарахгүй нь тодорхой байв. Энэ асуудлыг шийдвэрлэхийн тулд дүрэмд суурилсан, өгөгдөл жишээнд суурилсан, хосолсон зэрэг орчуулгын олон янзын програм туршиж байснаас хэл шинжлэлийн үүднээс хэл тус бүрийн онцлогийг тодорхойлох, өвөрмөц хэлцийг орчуулах, орчуулгын явцад гарах алдааг тусгаарлах зэрэг үйлдлийн систем бүхий статистик болон электрон нерв горимын машины орчуулгын програмыг улам бүр хөгжүүлж байна.

Статистик машины орчуулга нь Канад болон Европын парламентын Канадын Хансард Корус түүвэр буюу англи-франц хэл дээрх баримт бичигт тулгуурлан статистик аргаар орчуулах аргачлал бөгөөд дээр дурдсан хэлнээ төстэй агуулга бүхий эхийг орчуулахад үр дүн сайтай байдаг байна. Гэвч бодит байдал дээр олон хэлнээ иймэрхүү хос хэлний баримт бичиг тохиолдох нь ховор юм. 2005 онд *Google* корпораци орчуулгын програмаа сайжруулах зорилгоор НҮБ-ын баримт бичгээс 200 тэрбум орчим үгийг авч ашигласан байдаг.

Тодорхой даалгавар бүхий алгоритмаас илүүтэйгүүр гүн бүтцэд тулгуурлан өөрөө суралцах электрон нерв горимын машины орчуулга нь сүүлийн жилүүдэд эрчимтэй хөгжиж байна. *Google* корпорацийн орчуулгын програм нь эдүгээ өмнөх

статистик аргаас илүүтэйгээр энэ технологид суурилж байна. Мөн 2017 онд машины орчуулгын салбарт идэвхтэй үйл ажиллагаа явуулдаг *Kantan MT, Omniscien Technologies*, болон *SDL* зэрэг томоохон компаниуд электрон нерв горимын машины орчуулгын системд шилжиж буйгаа зарласан юм. Электрон нерв горимын машины орчуулгын давуу тал нь орчуулгын эхийг (*source text*) орчуулах хэлний (*target language*) өгүүлбэрзүйн бүтцээр задалж, задалсан өгүүлбэрийн бүтцийг хялбаршуулах, хялбаршуулсан өгүүлбэрийг эх хэлний өгүүлбэрийн бүтэц рүү буцаан хөрвүүлэх, тухайн хэлний ямар нөхцөл, дагавар залгахыг тодорхойлох, өгүүлбэрт илүүц үгийг хассаны үндсэн дээр хоёр хэлний харгалзах утгатай үгсээр орлуулан, эцэст нь сонгогдсон нөхцөл дагаврыг зөв бичгийн дүрмийн дагуу залгаж орчуулгыг дуусгадаг байна. Иймд орчуулгын сургалтад техникийн тусгай хэл, найруулгын нэр томъёо, байнга хэрэглэгддэг хэлзүйн дүрэм, үгсийн сан, өгүүлбэрийн бүтэц, найруулгын онцлог зэргийг анхаарах хэрэгтэй. Нөгөөтэйгүүр энэ чиглэлийн хэл-найруулгын онцлог, нэр томъёо, өгүүлбэрийн бүтцийг сайтар судалж, орчуулгын хичээлийн агуулгыг баяжуулах, оюутны эрдэм шинжилгээний ажлыг энэ чиглэлд хөтлөх нь ирээдүйн сайн орчуулагч, сайн орчуулгын програм хангамж бэлдэхэд үлэмжхэн хувь нэмэртэй болов уу.

Орчуулгын програм хангамж цаг ямагт өмнөх орчуулгын түүхийг хадгалж, тэр хэрээр орчуулгын чанар нь дээшилсээр байна. Орчуулгын програм хангамж нь техникийн орчуулгын үед нэр томъёог оновчтой оноогоод зогсохгүй хэлзүй, найруулгазүйн хувьд оновчтой байхаар програмчилагдаж, орчуулах болон хэвлэх үйл ажиллагааг ч хүртэл нэгэн зэрэг хийдэг болсон байна. Техникийн гарын авлага болон бусад баримт бичгийн агуулга шинэчлэгдэж байдаг хэдий ч дийлэнх хэсэг нь агуулгын хувьд төдийлөн өөрчлөгддөггүй бөгөөд орчуулгын програм санах ойн тусламжтайгаар өмнө нь орчуулж байсан хэсгийн байршлыг тогтоож, түүнийг дахин ашигладаг байна. Гэвч машины орчуулгын системийг боловсруулах нь зардал ихтэйгээс гадна тухайн салбар орчуулах хэл бүр дээр ижил түвшний, хангалттай тооны нэр томъёоны өгөгдлийн сантай болсон байх зэрэг нарийн шаардлагууд бас байдаг.

Сүүлийн үед орчуулгын програм хангамж улам бүр боловсронгуй болж, англи, франц, герман зэрэг тодорхой хэлний хүрээнд хувийн компьютерт зориулсан хувилбарыг ч мөн зах зээлд нийлүүлээд байна. Үүнээс гадна интернет орчинд ашигладаг орчуулгын програмын хувилбар хэдийнээ та бидний хэрэглээ болоод удаж байгаа бөгөөд үүнийг дагаад онлайн орчуулгыг хянах үйлчилгээ явуулдаг бизнесийн компаниуд нэмэгдэж байна.

Дүгнэлт

Хэдийгээр машины орчуулгад шинэ дэвшилт гарсаар байгаа ч бие даасан мэргэжлийн орчуулгын програм бий болох харахан болоогүй байна. Утга зохиол, хууль эрхзүй гэх мэт хүний орчуулгын ур чадварын үр дүнд бий болох, утгыг нарийвчлан задлах шаардлагатай эх материалын орчуулгын үед хүний орчуулга нэгдүгээрт хямд, хоёрдугаар илүү үр дүнтэй байх тул энэ төрлийн орчуулгын эрэлт

хэрэгцээ байсаар байх болно. Мөн аман орчуулгын салбарт машины орчуулга хүний орчуулгыг гүйцэх болоогүй байна. Дипломат болон бизнес яриа хэлэлцээрийн орчуулгад орчуулагчийг машинаар орлуулах боломж харагдахгүй байна.

Нөгөөтэйгүүр интернет дэх хэмжээлшгүй их цахим эх бичвэрийн орчуулгын хувьд хэдий орчуулгын чанар хангалтгүй ч хурдны хувьд машинтай өрсөлдөх зүйл алга л байна. Машины орчуулга нь хэмжээний хувьд хэт их, үг хэллэг, нэр томъёо, өгүүлбэрийн бүтцийн хувьд нэгэн жигд, дахин давтагдсан шинж чанартай, нэгэн зэрэг олон хэлэнд орчуулах техникийн баримт бичиг, бараа бүтээгдэхүүний гарын авлага зэрэг материалын орчуулгад оновчтой хувилбар байх нь тодорхой байна. Энэ мэтээр хүний болон машины орчуулга өөр өөрийн давуу болон сул талтай.

Бизнес захидал, харилцааны орчуулгад орчуулагч хүний үүрэг чухал хэвээр байх боловч хувийн шинжтэй захидал, харилцаанд машины орчуулгын хэрэглээ улам бүр нэмэгдэх хандлагатай байна. Үүнээс гадна электрон шуудан, вэб хуудас болон компьютерт суурилсан мэдээллийн үйлчилгээний салбарт машины орчуулга цорын ганц боломжит шийдэл нь байх болов уу.

Дүгнэж хэлэхэд машины орчуулгын үр дүнд орчуулгын салбарт шинэ шинэ боломжууд бий болсоор байна. Хэдийгээр машины орчуулга хүний орчуулгын түвшинд хүрэх болоогүй ч цаашид улам бүр боловсронгуй болох нь илэрхий. Нөгөөтэйгүүр сүүлийн үеийн орчуулгын програм муухан орчуулагчаас илүүтэй орчуулаад байгаа бодит байдлыг бид юу юунаас илүүтэй анхаарч, сургалтын хөтөлбөр, хичээлийн агуулга, заах аргазүйгээ шинэчилж байх нь орчуулгын чиглэлээр сургалт, судалгаа явуулж буй хэн бүхний хувьд чухал билээ.

Ашигласан эх сурвалж

1. https://en.wikipedia.org/wiki/Google_Translate#cite_note-GT_Turovsky_2016-2
2. Barak Turovsky, Found in translation: More accurate, fluent sentences in Google Translate, 2016
<https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>
3. John Hutchins, The development and use of machine translation systems and computer-based translation tools, 1999, Beijing, China
<http://hutchinsweb.me.uk/Beijing-1999.pdf>
4. Wheeler, Peter J. "Systran". In: Machine Translation Today: The State of the Art. Ed. by Margaret King. Edinburgh: Edinburgh University Press, 1987
5. Elliston, J.S.G. (1979) "Computer-aided translation – a business viewpoint." Translating and the Computer: proceedings of a seminar, London, 1978
6. Англи-Монгол орчуулгын онлайн системийн танилцуулга, 2010
<https://www.slideshare.net/jamuh/ss-6005596>