

МОНГОЛ УЛСЫН ИХ СУРГУУЛЬ
ШИНЖЛЭХ УХААНЫ СУРГУУЛЬ
“МОНГОЛ СУДЛАЛ”
Эрдэм шинжилгээний сэтгүүл

Боть XLVII (597)

DOI: 10.22353/ms20234716

МОНГОЛ ХЭЛ ЗҮЙН БҮТЭЭЛ ДЭХ ХЭЛ ШИНЖЛЭЛИЙН НЭР
ТОМЬЁОНЫ ХӨМРӨГИЙН БАЙГУУЛАЛТ*

Э.Уранбилэг
Давхарбаяр

Хураангуй: Хэлний материалын хөмрөг (corpus) нь нэгэн том хэлбэрийн мэдлэгийн сан болсны хувьд, хэл шинжлэлээс гадна нийгэм, амьдрал, улс төр, аж ахуй зэрэг олон талын асуудлыг судлахад хэрэгцээтэй мэдээлэл, тоо баримтаар хангаж өгдөг. Нэг хэлний их хэмжээний бичвэрийг цуглуулж, компьютерт хадгалан хөмрөгийг бүрдүүлдэг. Ингэж бүрдүүлсэн хөмрөгт элбэг баялаг хэлний мэдлэг нуугдаж байдаг бөгөөд түүнийг уудлан гаргаж бодит хэрэглээнд оруулахын тулд хөмрөгийг боловсруулах хэрэгтэй. Хөмрөгийг боловсруулснаар судалгаа ба хэрэглээнд ашиг тустай мэдлэгийг шүүн олж болно. Тиймээс судлаачид янз бүрийн хөмрөг бүрдүүлж, түүнээсээ өөрийн судалгаанд хэрэгцээтэй мэдээллийг олох зорилгоор нэг бус олон шат давхаргад нарийн задлалт хийж, тэмдэглэгээ хаддаг. Ингэж хөмрөгийг үг зүй, өгүүлбэр зүй, утга зүйн гэх мэт олон түвшинд боловсруулснаар толь бичиг зохиох, машин орчуулга хийх, өгүүлбэрийг автоматаар үүсгэх, авиа нийлүүлэх, мэдээлэл хайлт, автоматаар ялган таниулах зэргээр судалгаанд бүрэн ашиглах боломжтой юм. Энэхүү өгүүлэлд “Монгол хэлний зүйн нэр томьёог компьютерт автоматаар ялган таниулах” судалгааны хүрээнд байгуулж буй Б.Ринчен “Монгол бичгийн хэлний зүй” дөрвөн боть бүтээлд тулгуурласан бичгийн хэлний материалын хөмрөгийн бүрдүүлэлт, боловсруулалтын тухай өгүүлэх бөгөөд тухайн бүтээл дэх хэл шинжлэлийн нэр томьёоны салбар хөмрөг байгуулсан судалгааны үр дүнг танилцуулсан болно.

Түлхүүр үг: хэлний материалын хөмрөг, хөмрөг хэл шинжлэл, тэмдэг хадах, үг зүйн тэмдэг хадах, үг зүйн боловсруулалт, хэл шинжлэлийн нэр томьёо

1. Удиртгал

1.1. Хэлний материалын хөмрөгийн тухай

Нэг хэлний их хэмжээний бичвэрийг цуглуулж, компьютерт хадгалан хэлний материалын хөмрөг (corpus)-ийг бүрдүүлдэг. Үүнийг “боловсруулаагүй хөмрөг” буюу түүхий материалын хөмрөг гэж нэрлэдэг. Энэ нь нэг хэлний үнэн бодитой хэллэгийг тусгасан үлгэр хэв болж үг үсгийн давтамжийг бүртгэх, үгсийн хэлхээсийг гаргах, электрон толь бичиг найруулах зэрэг судалгаанд өргөнөөр ашиглагддаг байна.

* Тус судалгаа нь БНХАУ-ын нийгмийн шинжлэх ухааны суурь мөнгөний голдлогот төсөл “Боловсруулсан хэлний материалын хөмрөгт суурилсан монгол хэлний өгүүлбэр зүйн хэмжээ судалгаа” (зөвшөөрсөн номер: 19AYY018) ба Өвөр Монголын их дээд сургуулийн хэрэглээний хэл шинжлэлийн шинийг үүдэх багийн төслийн дэмжлэгийг хүртэв.

Хэлний материалын хөмрөгт хэлний баялаг мэдлэг нуугдаж байдаг бөгөөд түүнийг уудлан гаргаж бодит хэрэглээнд оруулахын тулд хэлний материалыг боловсруулах хэрэгтэй юм. Судлаачид хэлний материалын янз бүрийн хөмрөгөөс хэрэгцээтэй мэдээллээ олох зорилгоор олон шатлалтай нарийн задлалт хийж, тэмдэглэгээ хаддаг. Иймэрхүү хэлний материалын хөмрөгийг “боловсруулсан хөмрөг” гэнэ.

Хэлний материалын хөмрөгийн боловсруулалтад ярианы хэлний ба бичгийн хэлний материалын хөмрөгийн боловсруулалт гэж хоёр зүйл байх бөгөөд бичгийн хэлний материалын хөмрөгийн боловсруулалтад голдуу үг зүй, өгүүлбэр зүй, утга зүйн боловсруулалт хийхийг хэлдэг. Үүнд “үг зүйн боловсруулалт” нь хамгийн үндсэн зүйл болно.

Энэхүү өгүүлэлд Б.Ринчен “Монгол бичгийн хэлний зүй” (2015) дөрвөн боть бүтээлд тулгуурласан бичгийн хэлний материалын хөмрөгийн бүрдүүлэлт, боловсруулалтын тухай өгүүлэх бөгөөд түүний үр дүнд тухайн бүтээл дэх хэл шинжлэлийн нэр томъёоны салбар хөмрөг байгуулсан судалгааны дүнг танилцуулах юм.

1.2. Монгол хэлний зүйн бүтээлийн хөмрөгийн бүрдүүлэлт

Давхарбаяр нар “Хэлний материалын хөмрөг бол системт инженеринг болно. Түүнийг нээн ашиглахад, хэлний материалын хөмрөгийн төсөвлөлт, хэлний материалын хөмрөгийн цогцлолт, хэлний материалын хөмрөгийн боловсруулалт ба хэлний материалын хөмрөгийн хэрэглээ зэрэг явцыг дамждаг” (2018: 251) хэмээн хөмрөг байгуулах алхам дарааг дэлгэрүүлэн тайлбарласан байна. “Хөмрөгийн төсөвлөлт”¹ гэдэгт тухайн хөмрөг хэдэн салбар хөмрөгтэй байх, тэмдэглэх кодын загвар, тоо баримтын төрөл хэлбэр, хэмжигдэл авах хүрээ гэх мэт асуудал холбогдоно. “Хөмрөгийн цогцлолт”² гэдэгт голдуу хэлний материалыг сонгох ба гардан байгуулах зэргийг багтаана. Үүнд хөмрөгийг нэг загварт нэгдэлтэй тэмдэглэх, хадгалах бүтэц, харгалзсан хөмрөгийн хамааралтын систем, хайлтын системийг төлөвлөх зэрэг багтана.

Бидний судалгааны ерөнхий зорилго нь хэл зүйн нэр томъёог компьютерт автоматаар ялган таниулах учраас тэрхүү зорилгодоо үндэслэн монгол хэлний хэл зүйн бүтээлээс сонгон хөмрөг байгуулахаар төлөвлөсөн болно. Хэл шинжлэлийн ерөнхий онол, сургаалыг судалгааны үндэс, аргаа болгож монгол хэлний олон янзын үзэгдлийг тал бүрээс нь шинжлэн туурвисан гадаад, дотоодын эрдэмтдийн зохиосон хэл зүйн бүтээл олон буй. Тэдгээр бүтээлээс монгол хэлний бүтэц тогтолцоог нарийвчлан судалсан, орчин цагийн монгол хэлний онолын хэл зүйн бүтээл болох Б.Ринчен “Монгол бичгийн хэлний зүй” дөрвөн боть бүтээлийг сонгон авав.

Б.Ринчен Монгол улсын 39 онд буюу 1950-аад онд “Монгол хэлний зүйн бичиг” хэмээх суурь судалгааны бүтээлээ уламжлалт монгол бичгээр бичиж хэвлүүлсэн байдаг. Энэхүү бүтээлийн нэгэн хувь Унгар улсын ШУА-ийн номын санд хадгалагдаж байсныг хоёр улсын шинжлэх ухааны байгууллагуудын удирдлагын хүрээнд хэлэлцэн, Монгол Улсын ШУА-ийн Хэл зохиолын хүрээлэнд цахим хувилбарыг шилжүүлэн өгчээ. Уг бүтээлийг тус хүрээлэнгээс

¹ ӨМӨЗО-нд хэвлэгдсэн “Компьютер хэл шинжлэл”-ийн ном бүтээлд хэрэглэдэг нэр томъёо.

² Мөн адил.

кирилл үсэгт буулган 2015 онд “Монгол хэлний зүйн бичиг” (Удиртгал, авиа, үг, өгүүлбэрийн зүй) нэртэйгээр хэвлэсэн байна. Улмаар дээрх бүтээлээ дэлгэрүүлэн “Монгол бичгийн хэлний зүй” хэмээх 1100 гаруй хуудастай ганц сэдэвт зохиол бичиж 1956 оны 8 сард Мажар улсын ШУА-д докторын зэрэг хамгаалжээ. Үүнээс хойш нэлээд хэдэн жилийн дараа “Монгол бичгийн хэлний зүй” гэдэг иж бүтээлийнх нь тэргүүн дэвтэр “Удиртгал” 1964 онд, дэд дэвтэр “Авиан зүй” 1966 онд, гутгаар дэвтэр “Хэлбэр судлал” 1966 онд, дөтгөөр дэвтэр “Өгүүлбэр зүй” 1967 онд кирилл үсгээр хэвлэгдсэн билээ. Зохиогч уг бүтээлээ монгол бичгээр туурвисан боловч “... хэвлэлд шилжүүлэх үес кирилл үсгээр хэвлэвэл дээр гэсэн зөвлөмжийн дагуу монгол үсгээс нь кирилл үсэгт буулгах ажлыг өөрөө бичгийн машинаар цохиж, урьд бичсэн эх нооргийг үг найруулгыг сийрүүлэхийн далимд бас дээрдүүлж үзье гэж хэвлэлийн эхийг кирилл латин хоёр үсгийн машинаар цохиж гүйцээгээд, уйгуржин монгол, тод монгол үсгээр бичсэнийг нөхөн бичингүүт хянан гүйцээлээ” (Ринчен, 2015: 68) хэмээн тэргүүн дэвтрийнхээ оршилд өгүүлсэн байдаг.

1964-1967 онд хэвлэгдсэн дээрх дөрвөн дэвтэр бүтээлийн уг хэвлэл нь одоо олдоц муу болсон бөгөөд түүнээс хойш Монгол улсад болон ӨМӨЗО-д хэд хэдэн удаа хэвлэгдсэн байна. Тухайлбал, 2012 онд Монгол Улсын Шинжлэх Ухааны Академийн Хэл зохиолын хүрээлэнгээс эрхлэн гаргасан “Бямбын РИНЧЕН. Эрдэм шинжилгээний бүтээлийн чуулган” хэмээх цувралын V, VI, VII, VIII боть; 2015 онд Боловсрол, Соёл, Шинжлэх ухааны яам болон Монгол судлалын үндэсний зөвлөлийн эрхлэн хэвлүүлсэн “Монгол судлалын шилдэг өв” цувралаар Б.Ринчений эрдэм шинжилгээний бүтээлийн чуулган 9 боть бүтээлийн эхний I, II, III, IV ботид тус тус хэвлэгдсэн байна. ӨМӨЗО-ны Өвөр монголын ардын хэвлэлийн хорооноос 1991 онд тэргүүн дэвтрийг, 1992 онд дэд, гутгаар, дөтгөөр дэвтрийг тус тус монгол бичгээр хөрвүүлэн хэвлэжээ. Үүний дараа 2016 онд хоёр дахь хэвлэлийг хэвлэж, хамгийн сүүлд 2020 онд гурав дахь хэвлэлийг хэвлэн тархаасан байна.

Хэлний материалын хөмрөгийг бүрдүүлэхийн тулд анх цаасан хэвлэмэл бүтээлээс хүний хүчээр шивж компьютерт файл болгон хадгалах замаар нэлээд цаг зарцуулан бүтээж байв. Үүний дараа зураг буюу дүрсийг компьютерт таниулж боловсруулах технологи хөгжиж, цаасан эхийг сканердан зураг болгоод оруулахад бичвэрийг файл болгодог программ гарчээ. Энэ нь хөдөлмөрийг хөнгөвчлөх боловч тэмдэгтийг бүрэн зөв танихгүй учраас тухайн уг эхтэй нь дахин тулгаж засварлах шаардлагатай байдаг. Мөн “цахим орчноос хөмрөг бүрдүүлэхийн тулд янз бүрийн хөтөч, медиа платформуудтай зохицон ажилладаг медиа мониторингийн систем хэрэглэсэн туршлага гарч байна” (Мөнх-Учрал, 2020).

Бид энэхүү судалгаандаа Монгол Улсад 2015 онд “Монгол судлалын шилдэг өв” цувралаар хэвлэгдсэн 9 ботийн эхний 4 боть кирилл үсгээр хэвлэгдсэн номын эхийг гол болгон баримталж, 2020 онд ӨМӨЗО-д хэвлэгдсэн монгол бичгээрх эхтэй харьцуулан үзэж үндсэн эх хэрэглэгдэхүүн болгон ашигласан болно. Хөмрөгийг бүрдүүлэхдээ эхлээд кирилл үсгийн эхийг OCR мэргэжлээр³

³ OCR - Optical Character Recognition, OCR Print – Хэвлэмэл оптик тэмдэгт таних систем ашиглан тэкст болгохыг хэлнэ. Бид номын сканнердсан эхийг ABBYY FineReader 15.0 программ ашиглан хөрвүүлэв.

дамжуулан текст файл болгов. Түүний дараа үндсэн номын эхтэй тулган засав. Дараа нь кирилл бичгээс монгол бичигт хөрвүүлдэг хөрвүүлэгч программ⁴ ашиглан монгол бичигт хөрвүүлэв. Монгол бичигт хөрвүүлсэн эхээ 2020 оны ӨМӨЗО-нд хэвлэгдсэн дөрвөн дэвтэр монгол бичгээрх эхтэй тулган харгуулж буруу хөрвүүлснийг засав. Эдгээр цаг хугацаа, хүч хөдөлмөр шаардсан ажлын үр дүнд Б.Ринчен “Монгол бичгийн хэлний зүй” 4 боть бүтээлийн кирилл болон монгол бичгээрх боловсруулаагүй хэлний материалын хөмрөг бэлэн болов.

2. Судалгааны хэсэг

2.1. Монгол хэлний зүйн бүтээлийн хөмрөгийн боловсруулалт

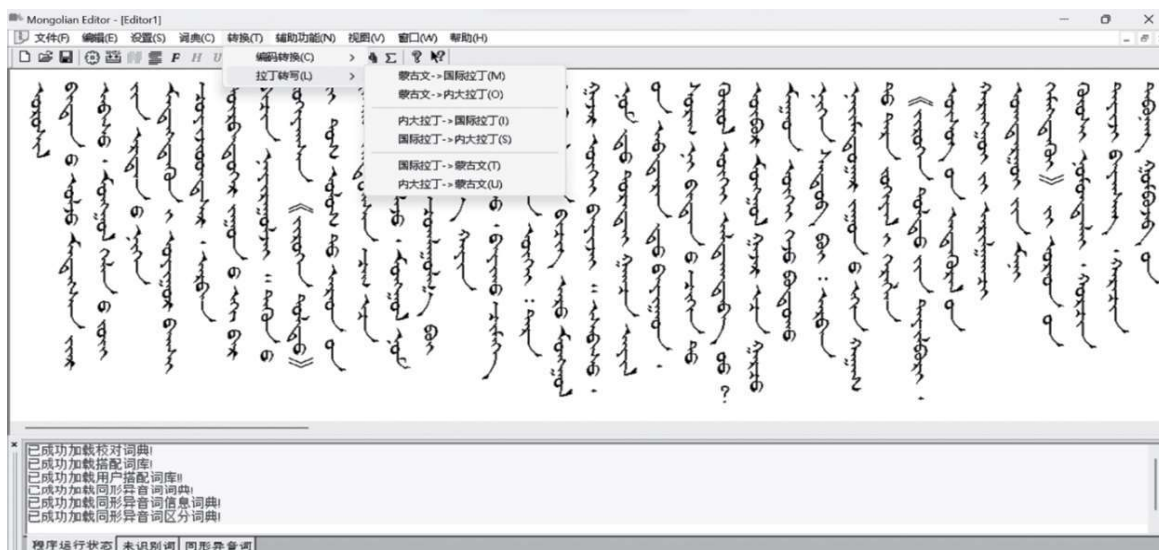
Өмнөх хэсэгт өгүүлсэн монгол бичгээрх боловсруулаагүй хөмрөг бол бидний үндсэн эх материал болно. Уг материалыг программын тусламжтай боловсруулж латин галигт буулган, улмаар үг зүйн тэмдэг хадах нь бидний судалгааны зорилго юм. Учир нь энэхүү судалгаа нь “Монгол хэлний зүйн нэр томъёог компьютерт ялган таниулах” судалгааны нэг хэсэг болно. Үүний тулд бид хөмрөгийн боловсруулалт хийхдээ уг эх материалаа латин галиг болгож, түүний дараа үгсийн аймаг, үгийн хувиллын тэмдэг хадсанаар компьютерт таниулах нөхцөл болон дүрмийг хураангуйлж гаргах боломжтой болох юм. Энэ хэсэгт хэлний материалын хөмрөгийн үг зүйн боловсруулалтыг хэрхэн хийсэн алхам дараа болон ашигласан программын тухай дэлгэрүүлж өгүүлье.

2.1.1. Хэлний материалын хөмрөгийг латин галигт буулгасан нь

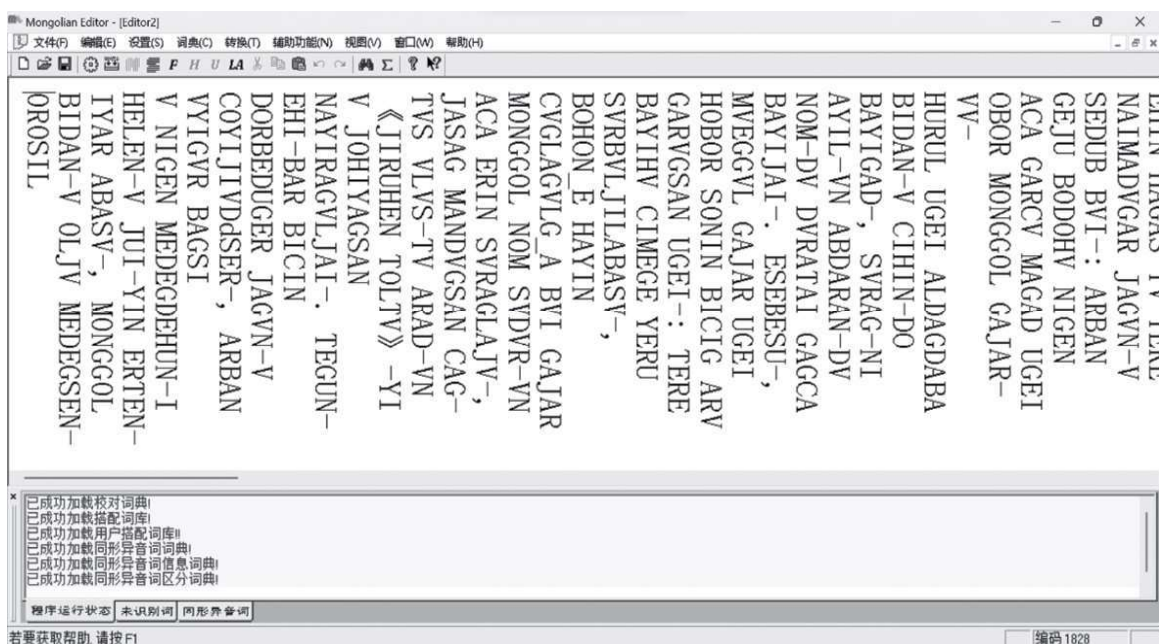
Монгол хэлний зүйн бүтээлийн монгол бичгээрх эх материалыг латин галигт буулгахын тулд монгол хэл бичгийн материалыг боловсруулахад зориулан бүтээсэн тусгай программ ашиглав. Өвөр Монголын Их Сургуулийн Монгол судлалын дээд сургуулийн багш, судлаачид 2006 оноос эхлэн үгийн зөв бичлэгийн алдаа засах, үгийн хувиллын тэмдэг хадах систем боловсруулж эхэлсэн байна. Улмаар түүнийгээ шинэчлэн сайжруулсаар 2012 онд монгол бичгээрх бичвэрийг боловсруулдаг Mongolian Editor for Plain Text 4.0-ийг үйлдэн гаргав. 2017 онд энэхүү системээ халх, буриад, халимаг зэрэгт тархаан хэрэглүүлэх үүднээс олон хэлээр ажиллах боломжтой болгон өөрчилж хэвлэлийн дугаараа 5.0 болгон нэмэгдүүлжээ. Энэхүү “Монгол бичгийн цул текст харгуулгын систем ME (Mongolian Editor)”⁵ нь монгол бичгээрх бичиг материалын алдааг засах, бичиг материалыг найруулах, код хөрвүүлэх, латин галигаар буулгах, толгой үгсийн жагсаалт хийх, үеэр салгах, үгсийн бүртгэл хийх гэх мэт монгол хэл боловсруулалтын олон талын үйлдэл чадамжтай хэрэглээний программ юм. Энэхүү программыг ашиглан Б.Ринчен “Монгол бичгийн хэлний зүй” 4 боть бүтээлийн монгол бичгээрх хөмрөгийг латин галигт буулгав.

⁴ <http://trans.mglip.com/CyrillicC2T.aspx> хөрвүүлэгч болон ӨМӨЗО-ны “Мөнх гал” компанийн тусламжтай хөрвүүлэв.

⁵ 斯。劳格劳 Mongolian Editor for Plain Text Version 5.0. 2018



Зураг 1. Mongolian Editor программд монгол бичгийн хөмрөгийг оруулсан байдал



Зураг 2. Mongolian Editor программаар латин галигт буулгасан байдал

Жишээ болгон “Монгол бичгийн хэлний зүй” тэргүүн дэвтэр “Удиртгал” ботийн Оршил хэсгийн эхний өгүүлбэрийг кирилл бичгээс монгол бичиг, латин галиг болгосныг үзүүлбэл дараах байдалтай болно.

Бидний олж мэдсэнээр аваас, монгол хэлний зүйн эртний нэгэн мэдэгдэхүүнийг уйгур багш Чойжи-одсэр, арван дөрөвдүгээр зууны эхээр бичин найруулжээ.

ᠪᠢᠳᠡᠨᠤ ᠣᠯᠵᠢᠪᠡᠭᠰᠡᠨᠢᠶᠡᠷ ᠠᠪᠠᠰᠤᠪᠠ, ᠮᠣᠩᠭᠣᠯ ᠬᠡᠯᠡᠨᠤ ᠵᠢᠤᠶᠢᠨ ᠡᠷᠲᠡᠨᠤ ᠨᠢᠭᠡᠨ ᠮᠡᠳᠡᠭᠳᠡᠬᠡᠳᠡᠬᠡᠨᠤᠨᠢ ᠠᠶᠢᠭᠦᠷ ᠪᠠᠭᠰᠢ ᠴᠣᠶᠢᠵᠢᠠᠳᠰᠠᠷ, ᠠᠷᠪᠠᠨ ᠳᠣᠷᠪᠡᠳᠦᠭᠦᠷ ᠵᠠᠭᠦᠨᠤᠪᠤ ᠡᠬᠢᠪᠦᠷ ᠪᠢᠴᠢᠨ ᠨᠠᠶᠢᠷᠠᠭᠦᠯᠵᠠᠢ.

BIDEN-U 0LJV MEDEGSEN-IYER ABASV, MONGGOL HELEN-U JUI-YIN ERTEN-U NIGEN MEDEGDEHUN-I VYIGVR BAGSI C0YIJ10DSAR, ARBAN DORBEDUGER JAGVN-V EHI-BER BICIN NAYIRAGVLJAI.

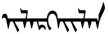
Хэлний материалыг боловсруулахад, хэлний материалын хөмрөгийн үг нэг бүрийн бичгийн дуудлагыг үндэслэж бүрэн зөв латинчлан галигласан байх хэрэгтэй. Тиймээс энэхүү латин галигийн хөмрөгөө “Одоо үеийн монгол хэлний материалын хөмрөгийн галиглах дүрэм”⁶-ийн ёсоор хянан шалгаж буруу галигласан зүйлсийг 2020 онд Өвөр монголын ардын хэвлэлийн хорооноос хэвлэсэн “Монгол бичгийн хэлний зүй” дөрвөн дэвтэр номтой тулган залруулав. Тус хэлний материалд үгийг буруу галигласан байдлыг үзүүлбэл дараах хэдэн байдлаар буруу зүйл гарсан байна. Үүнд,

1. Монгол бичгээр дүрс ижил боловч дуудлага өөр үсэг хоорондоо солигдож латин галиг нь буруу гарсан тохиолдол олон байв. Жишээ нь:

–  хэмээх үгийг буруу галигласан нь VRVSIL, зөв хувилбар нь 0R0SIL юм. Учир нь “Одоо үеийн монгол хэлний материалын хөмрөгийн галиглах дүрэм”-д эгшгийг дараах байдлаар тэмдэглэдэг. О эгшгийг 0, У эгшгийг V гэх мэт.

	A	E	I	0	V	O	U
	а	э	и	о	у	о	у
	а	э	и	о	у	о	у
	а	э	и	о	у	о	у

–  хэмээх үгийг буруу галигласан нь BIDAN-V, зассан нь BIDEN-U;

–  хэмээх үгийг буруу галигласан нь ILERGEYILEN, зассан нь ILERHEYILEN гэх мэт.

2. Монгол бичгээрх эхэд үгийг буруу бичсэнээс латин галиг буруу тэмдэглэгдсэн тохиолдол байв. Жишээ нь: DoirelkEt буруу нь TURILGE, зассан нь TURELGE,  буруу нь JARAG-VN SEGUL, зассан нь JARGV-YIN SEGUL гэх мэт.

3. Бусад шалтгаанаар буруу тэмдэглэгдсэн зүйлс. Жишээ нь:  буруу нь NAYIRAGVLGA, зассан нь NAYIRAGVLG_A,  буруу нь BICIG, зассан нь BICIG гэх мэт.

Энэ мэтээр программ ашиглан латин галигт буулгахад алдаатай галигласан үгсийг хүний хүчээр засаж залруулсны дүнд монгол хэлний зүйн бүтээлийн латинаар галигласан хөмрөг бэлэн болов. Уг хөмрөг нь .txt өргөтгөлөөр хадгалагдана.

2.1.2. Хэлний материалын хөмрөгт үг зүйн тэмдэг хадсан нь

Хэлний материалын хөмрөгийн үг зүйн боловсруулалтад үг таслах, үгсийн аймгийн тэмдэг хадах, тусгай нэрсийн тэмдэг хадах зэрэг агуулга багтдаг. Үг таслах гэдэг нь компьютер толь бичгээр үндэс болгож хэлний материалыг үгээр нэгж болгон таслахыг хэлнэ. Үгсийн аймгийн тэмдэг хадах гэдэг нь хэл судлаачид ямар нэгэн хэл шинжлэлийн онолд суурилж, судалгааны хэрэгцээгээ

⁶ Өвөр монголын их сургуулийн Хэл судлах газрын боловсруулсан монгол бичгийн хэлний материалын хөмрөгийг латинаар галиглахад журамлах дүрэм

үндэслэн хөмрөгийн доторх үг нэг бүрд үг зүйн мэдээллийг хадахыг хэлнэ. Жишээлбэл, нэр үгийн ганц ба олон тоо, тэмдэг нэрийн зэрэг, үйл үгийн төлөв гэх мэт. Анх 1971 онд Brown их сургуулийн судлаачид, Brown хөмрөгт үгсийн аймгийн тэмдэг хадсан бөгөөд тэд нийт 77 тэмдгийг тогтоож, тус тэмдэг хадах программын оновчийн хэмжээ нь 77 хувьд хүрч, хүний хүчээр зассаны дараа нэлээд өндөр хэмжээнд хүрч, дэлхийн хамгийн анхны үг зүйн боловсруулалт хийсэн хэлний материалын хөмрөг болжээ.

Хэлний материалын хөмрөгт үг зүйн боловсруулалтыг хийхдээ дүрэмд суурилсан, бүртгэлд суурилсан, дүрэм ба бүртгэлд суурилсан зэрэг аргаар хийнэ. Дүрэмд суурилсан арга гэдэг нь хэл шинжээчид тухайн хэлний тухай мэдлэгээ ашиглан хураангуйлж гаргасан, хэлбэрчлэн илтгэсэн дүрмүүдийг ашиглаж, хэлний материалын хөмрөгт үгсийн аймгийн тэмдэг хадахыг хэлнэ. Жишээ нь, “Brown хөмрөгийн TAGGIT систем нь дүрэмд суурилсан аргыг хэрэглэж хөмрөгт тэмдэг хаддаг систем юм. Тус систем нь 3300 зурвас дүрэм ба 86 зүйлийн үгсийн аймгийн тэмдгийг хэрэглэж, оновчийн хэмжээ нь 77 хувьд” (Wen Feng, 2020) хүрчээ.

Монгол хэл нь залгамал хэлэнд багтаж, үгсийн хувиллаар маш баялаг байж, үгийн язгуур үндсэнд дагавар залгахдаа үгийн араас дагуулж болон нийлүүлж залгадаг. Иймээс монгол хэлний материалын хөмрөгийн үг зүйн боловсруулалтын тэргүүн үүрэг бол монгол хэлний үгсийн үндэс ба үг хувилгах нөхцөл, дагаврыг нэг нэгээр нь тасалж сэргээх явдал юм. Өвөр монголын их сургууль “Одоо үеийн монгол хэл бичгийн материалын хөмрөгийг цогцлон байгуулсны дараа, монгол хэлний дагаврыг компьютероор автоматаар таслах шаардлага буй болжээ. Энэхүү хэрэгцээнээс үүдэж, хэдэн зүйлийн толь бичиг, хэл зүйн нөхцөлийг найруулж, тусгай дүрэм зохиож, монгол хэлний үгийн язгуур, үндэс, дагаврыг автоматаар таслах системийг боловсруулжээ” (Xiao Juan, 2016). Энэхүү системийн дагуу үгийг таслахад монгол хэлний тийн ялгал, олон тооны нөхцөлийн өмнө дундуур зураас << – >> -ыг тавьж, бусад хувилал болон үйл үгийн хувиллын өмнө жишүү зураас << / >> тавьж тэмдэглэжээ. Жишээ нь, *аавын* хэмээх үгийг тасалж *аав-ын* буюу ABV-YIN, *яваад* гэхийг *яв/аад* буюу YABV/GAD. Энэ нь монгол хэлний материалын хөмрөгийн үг зүйн боловсруулалтын эхлэл болжээ. 2004 оноос Өвөр монголын их сургуулийн Монгол хэл бичиг судлах газраас AYIMAG болон DARHAN системийг боловсруулан гаргажээ. Эдгээр систем нь дүрэм бас толь бичгийг уялдуулах аргыг хэрэглэж, оновчийн хэмжээ нь тус тус 83 ба 77.2 хувьд хүрсэн байна. 2010 онд “ӨМИС-ийн Монгол судлалын дээд сургууль ба БНХАУ-ын Шинжлэх ухааны хүрээлэнгийн Тооцоо мэргэжлийн судлах газар хамтран бүртгэлд суурилсан үгсийн аймгийн тэмдэг хадах систем Mglex-ийг гаргажээ. Тус систем нь монгол хэлний үгсийн хувиллын онцлогт суурилж, монгол хэлний үг зүйн задлалтад зориулсан тусгай бүртгэл загварыг зохиож гаргасан бөгөөд үг таслах оновчийн хэмжээ нь олон дахин сайжруулсны дүнд 97 хувьд хүрчээ.

Энэхүү судалгаанд монгол хэлний зүйн бүтээлийн хөмрөгт үг зүйн боловсруулалт хийхдээ Mglex системийг ашиглав. Латин галигт үгсийн аймгийн тэмдэг хадсаны дараа буруу хадсан тохиолдол гарах тул түүнийг хүний хүчээр засах шаардлагатай байдаг. Үүний тулд тусгайлан боловсруулсан үгсийн аймгийн тэмдгийн чуулган, стандартыг баримтлах ёстой. Бид программаар

тэмдэг хадсаныг хянаж залруулахдаа 2011 онд нийтэлсэн ӨМӨЗО-ны “Сураг зангийн мэргэжил – сураг занги шийдвэрлэлтэд зориулсан монгол хэлний үгсийн ангилал жич тэмдэг” (GB / T26235 – 2010) хэмээх стандартыг баримтлав. Энэхүү тэмдгийн чуулган нь үгсийн аймгийн тэмдэг, хувилгах дагаврын тэмдэг, бусад тэмдэг гэсэн гурван хэсгээс бүрдэнэ. Үгсийн аймгийн тэмдгийг гурван дэсээр хуваажээ. Үүнд, нэгдүгээр дэсэд жинхэнэ нэр N (noun), тэмдэг нэр A (adjective), тооны нэр M (numeral), хэмжүүрийн нэр Q (quantity), цаг орны нэр O (oron), төлөөний нэр R (pronoun), үйл үг V (verb), цагийн нэр T (time), дайвар үг D (adverb), хандлага үг H, дүрслэх үг U, дагавар үг G, сул үг S, холбох үг C, аялга үг I (interjection) гэж 15 төрөл зүйлд хуваасан байна. Эдгээр 15 төрөл дотроо дахин хуваагдана. Жишээ нь, жинхэнэ нэр дотроо энгийн нэр Ne, тусгай нэр Nt гэсэн хоёр төрөлд хуваагдаад энгийн нэр нь тоолдог нэр Ne1, эс тоолдог нэр Ne2, тусгай нэр нь хүний нэр Nt1, газрын нэр Nt2, байгууллагын нэр Nt3, нэр томъёо Nt4 гэсэн 4 төрөл болно. Хувилгах дагаврын тэмдэг нь 4 дэстэй. Нэгдүгээр дэс нь хувилгах дагавар F (suffix), хоёрдугаар дэс тооны ай Fp (plural), гуравдугаар дэс олон тооны дагавар Fp1, дөрөвдүгээр дэс олон тооны *нар*² дагавар Fp11 гэх мэтээр 196 зүйлийн тэмдэг тогтоожээ. Гуравдугаар хэсэг нь бусад тэмдэг буюу монгол тоо Et1, араб тоо Et2, хоршоо үг Y, цэг тэмдэг Wp гэх мэт дөрвийн дэсийн 43 төрлийн тэмдэгтийг жагсаасан байна.

Жишээ болгон монгол хэлний зүйн бүтээлийн хөмрөг дэх нэг өгүүлбэрт Mglex систем ашиглан үг зүйн тэмдэг хадсан байдлыг үзүүлбэл дараах байдалтай болно.

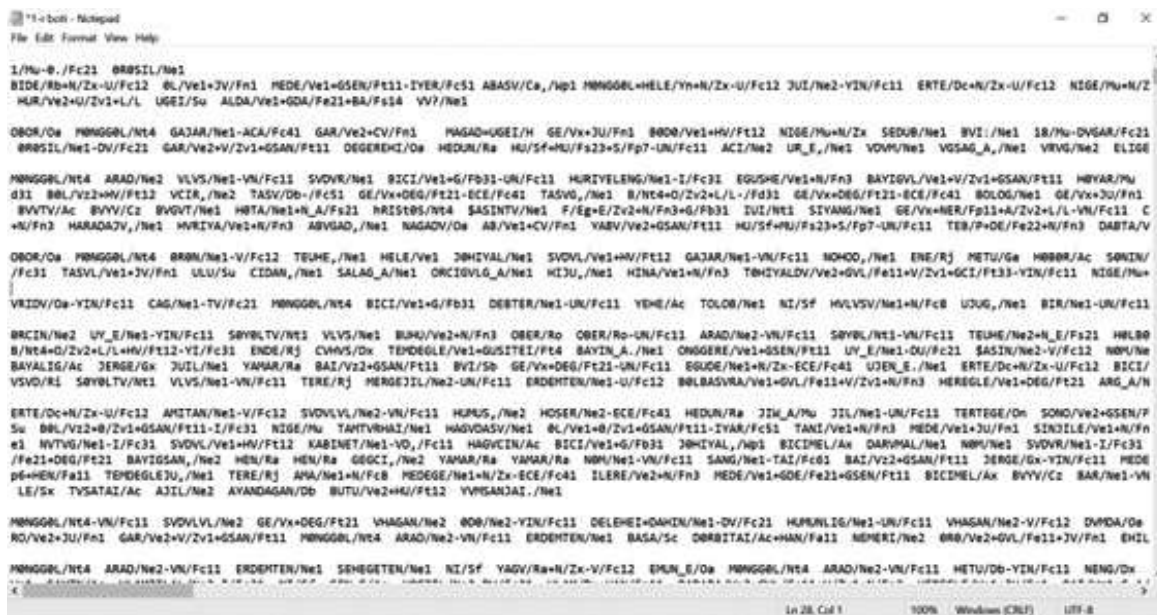
Тэмдэг хадаагүй латин галиг:

BIDEN-U 0LJV MEDEGSEN-IYER ABASV, M0NGG0L HELEN-U JUI-YIN
ERTEN-U NIGEN MEDEGDEHUN-I VYIGVR BAGSI C0YIJ0DSAR, ARBAN
DORBEDUGER JAGVN-V EHI-BER BICIN NAYIRAGVLJAI.

Үгсийн аймгийн тэмдэг, хувилгах дагаврын тэмдэг хадсан нь:

BIDE/Rb+N/Zx-U/Fc12 0L/Ve1+JV/Fn1 MEDE/Ve1+GSEN/Ft11-IYER/Fc51
ABASV/Ca,/Wp1 M0NGG0L=HELE/Yn+N/Zx-U/Fc12 JUI/Ne2-YIN/Fc11 ERTE/
Dc+N/Zx-U/Fc12 NIGE/Mu+N/Zx MEDEGDEHUN/Ne2-I/Fc31 VYIGVR/Nt4
BAGSI/Ne1 [C0YIJ0DSAR/Nt1,/Wp1 ARBAN=DORBEDUGER/Yn JAGVN/
Ne1-V/Fc12 EHI/Ne2-BER/Fc51 BICI/Ve1+N/Fn3 NAYIRA/Ve2+GVL/Fe11+JAI/
Fs11./Wp1

Монгол хэлний зүйн бүтээлийн латин галигийн хөмрөгт тэмдэг хадах Mglex систем ашиглан үгсийн аймгийн болон үгийн хувиллын тэмдэг хадав. Програмаар тэмдэг хадсан байдалд алдаатай зүйл гарсныг “GB / T26235 – 2010” стандартын дагуу хүний хүчээр хянаж залруулав. Хянан засах явцад үгийг таслахдаа буруу тасалсан, зарим нөхцөл дагаврыг буруу таньж хадсан зэрэг алдаа ихэвчлэн тохиолдов. Эдгээр ажлын үр дүнд Б.Ринчен “Монгол бичгийн хэлний зүй” 4 боть бүтээлийн хөмрөгийн үг зүйн боловсруулалт хийх ажил дуусаж монгол хэлний зүйн бүтээлийн “боловсруулсан хөмрөг” бэлэн болов.



Зураг 3. Монгол хэлний зүйн бүтээлийн хөмрөгт үг
зүйн боловсруулалт хийсэн байдал

2.2. Монгол хэлний зүйн бүтээл дэх хэл шинжлэлийн нэр томьёоны салбар хөмрөгийн байгуулалт

Хэлний материалын хөмрөгт боловсруулалт хийснээр хэрэглээнд орох боломжтой болдог. Хэрэглээнд орсны дараа ч хөмрөгийг тасралтгүй бүтээн байгуулж, төгөлдөржүүлэх боломжтой. Тухайлбал, хэлний материалыг шинэчлэх, шинээр тэмдэг хадах, салбар хөмрөг байгуулах зэргээр хөмрөгийн чанар чансааг дээшлүүлж болно. Бид энэхүү хэсэгт үг зүйн боловсруулалт хийсэн монгол хэлний зүйн бүтээлийн хөмрөг дотроос хэл шинжлэлийн нэр томьёоны салбар хөмрөг байгуулсан явц, түүний онцлогийн тасалбар, хэмжигдлийн тухай өгүүлнэ. Үүний тулд эхлээд Б.Ринчен “Монгол бичгийн хэлний зүй” дөрвөн боть бүтээлээс үндсэн толгой үг буюу хэл шинжлэлийн нэр томьёог ялган түүвэрлэнэ. Нэр томьёог түүвэрлэхдээ тухайн үгийн энгийн үгээс ялгарах онцлог шинж буюу нэр томьёоны шинжийг илтгэж буй эсэх, хэл шинжлэлийн нэр томьёоны онцлогийг илэрхийлж буй эсэх, мөн хэл шинжлэлийн нэр томьёоны толь бичигт буй эсэх гэсэн үндсэн шалгуураар шүүн түүвэрлэхэд давхардаагүй тоогоор 600 гаруй нэр томьёо байв⁷.

Монгол хэлний зүйн бүтээл дэх хэл шинжлэлийн нэр томьёоны салбар хөмрөгийн үндсэн толгой үгийг кирилл, монгол бичиг болон латин галиг гэсэн гурван хэлбэрээр жагсааж, ард нь уртац, давтамж, үгсийн аймаг, үгсийн аймгийн бүрэлдэл, жишээ өгүүлбэр зэрэг мэдээллийг оруулав. Эдгээр мэдээлэл нь тус хөмрөгийн онцлогийн талбар (attribute field) болох бөгөөд тус бүрээр нь хэмжигдлийг тодорхойлбол дараах байдалтай болно.

⁷ Энэ тухай ШУА, ХЗХ-ийн “Хэрэглээний хэл шинжлэл”. Боть VIII (23), 182-199х. сэтгүүлд хэвлүүлсэн “Монгол хэл зүйн бүтээл дэх хэл шинжлэлийн нэр томьёоны шинжилгээ” өгүүлэлд дэлгэрэнгүй өгүүлсэн болно.

Ангилал. Толгой үгийг А-Я үсгийн дарааллаар жагсаан дотор нь хэл шинжлэлийн салбар ухааны дагуу зүйлчлэн ангилсан болно. Зүйлчлэлийг тэмдэглэхдээ HJ (HELE JUI) гэсэн үсгээр товчлон доторх ангиллыг HJ1-HJ6 хүртэл дугаарлав. Нийт 664 толгой үгс буйгаас HJ1 буюу үсэг зүйн нэр томьёо 151, HJ2 буюу авиа зүйн нэр томьёо 153, HJ3 буюу үг зүйн нэр томьёо 170, HJ4 буюу өгүүлбэр зүйн нэр томьёо 31, HJ5 буюу үгийн сангийн нэр томьёо 57, HJ6 буюу ерөнхий хэл шинжлэлийн болон бусад нэр томьёо 102 байна.

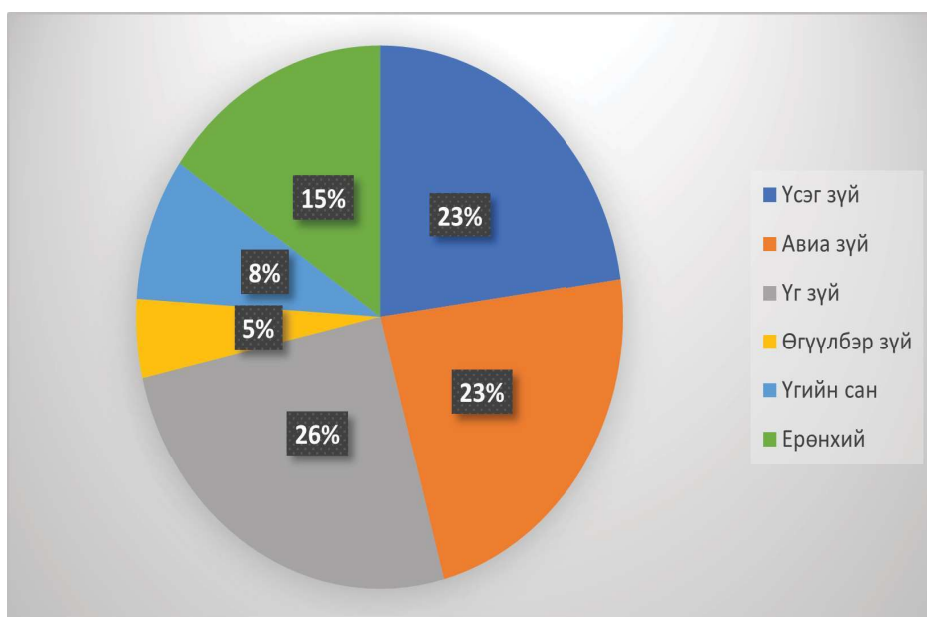


График 1. Хэлний зүйн бүтээл дэх хэл шинжлэлийн нэр томьёоны ангилал

Уртац. Уртац буюу толгой үгийг бүтээж буй үгийн тоо нь хамгийн бага нь буюу нэг үгээс бүтсэн 78, хоёр үгээс бүтсэн нь 339, гурван үгээс бүтсэн нь 210, дөрвөн үгээс бүтсэн 31, таван үгээс бүтсэн 5, зургаан үгээс бүтсэн 2 тохиолдол байна.

Үгсийн аймаг. Тухайн толгой үг ямар үгсийн аймагт харьяалагдаж буйг заана. Нийт толгой үгийн үгсийн аймгийн хувьд нэр үг хамгийн их буюу 600, үйл үг 43, цаг орны нэр 15, тэмдэг нэр 6 байна.

Үгсийн аймгийн бүрэлдэл. Үгсийн аймгийн бүрэлдэл гэдэг нь тухайн толгой үгийн уртац нь хоёр ба түүнээс дээш үгээс бүтсэн тохиолдолд тэдгээр үгс нь ямар үгсийн аймгаас бүрэлдэж буйг заана. Жишээ нь “авиа гээгдэх” хэмээх нэр томьёоны эхний үг нь нэр үг, хоёр дахь үг нь үйл үг тул N+V гэсэн үгсийн аймгийн бүрэлдэлтэй болох юм. Хоёр үгээс бүтсэн нэр томьёоны хувьд N+N *авианы үсэг*, A+N *бага үсэг*, M+N *ганц тоо*, O+N *доод өгүүлүүр*, Q+N *одоо цаг*, N+V *авиа ижиллэх*, V+N *болзох хэлбэр* гэсэн бүтэцтэй, гурван үгээс бүтсэн нэр томьёоны хувьд N+N+N *авиа зүйн галиг*, A+V+N *дан шүргэх авиа*, A+S+N *төрөл бус хэл*, A+N+N *амьд язгуур үг*, A+A+N *элдэв хоршоо үг*, M+N+N *ганц тоот*

ай, N+O+N хэлний урдуурх эгшиг, N+V+N биеийн заасан нэр, O+V+N ирээдүйн үргэлжилсэн цаг гэх мэт бүтцээр илэрчээ. Дөрвөн үгтэй нэр томьёо Q+V+V+N ойрдоо хүсэн болзох хэлбэр, таван үгтэй нь V+A+A+A+N үгүйсгэх эсрэг тэсрэг хоршоо үг гэх мэт байна.

Жишээ өгүүлбэр. Монгол хэлний зүйн бүтээлийн хөмрөгөөс тухайн толгой үг орсон өгүүлбэрийг хайж, хайлтад илэрч буй олон өгүүлбэр дундаас утгыг тодотгосон, товч бөгөөд тодорхой өгүүлбэрийг сонгон авч түүвэрлэсэн болно.

A	B	C	D	E	G	H	I
Король	Монгол бичиг	Латин гариг	Ангилал	Урсал	Үлэмж	Үл бүрэлдэл	Жишээ өгүүлбэр
ама	ᠠᠮᠠ	AMY A	HJ2	1	N	N	HELEN-U BUTUCE-YI SYDVLHV-DV, HOYAR JUL-UN UEGDEL GAI
ама бутн	ᠠᠮᠠᠪᠲᠤᠨ	AMY A BUTUHU	HJ2	2	V	N+V	EGUN-DU, AMYAN JU-YIN MEDEGDEHUN TVNG VCIR HOLBOGDA
ама гүндэг	ᠠᠮᠠᠭᠦᠨᠳᠡᠭ	AMY A GEGODEHU	HJ2	2	V	N+V	CTNGGRED MINGGOL HELETEN DVMIDA, HAMHIV AMY A NI GE
ама жажинг	ᠠᠮᠠᠵᠢᠨᠭ	AMY A JILJINGU	HJ2	2	V	N+V	MINGGOL HELETEN-U JARIM JARIM AYALGVN-DV MINGGOL UGE
ама сэлгн	ᠠᠮᠠᠰᠡᠯᠭᠢᠨ	AMY A SELGURU	HJ2	2	V	N+V	TERE HANGSIV GEDEG UGE-BER BIDOHV-DV, HENGSGU GED
ама тонгорол	ᠠᠮᠠᠲᠣᠨᠭᠣᠷᠠᠮᠤ	AMY A TONGGORAIV	HJ2	2	V	N+V	MAN-V HELEN-DU ODO CV INGGURU HEREGLEGDEU, BEY E-BEN I
амазбар	ᠠᠮᠠᠵᠠᠪᠠᠷ	AMFALABVRI	HJ2	1	N	N	EL E AMYAN-V JUL-ECE NGGE IERHEYEL-ECE NOGOGE-YI NI
амазбар сузал	ᠠᠮᠠᠵᠠᠪᠠᠷᠰᠠᠵᠠᠯ	AMFALABVRI SYDVLVL	HJ2	2	N	N+N	AGVSB-YIN HEI GARHV UYES DAGVN-V HOBCT DORGUV BAYIDA
амаа зүй	ᠠᠮᠠᠶᠠᠨᠵᠡᠢ	AMYAN JU	HJ2	2	N	N+N	HOGOLAI-YIN HEI GAGCA HU AMA-BAR DAMEN GARHV ABASY, I
амаа зүйн галг	ᠠᠮᠠᠶᠠᠨᠵᠡᠢᠭᠠᠯᠭ	AMYAN JU-YIN GALG	HJ2	2	N	N+N+N	TEGUN-I AMYAN-V JU-YIN GALG-TV, TEMDEG-IYER EGESG-UN
амааа үсг	ᠠᠮᠠᠶᠠᠨᠤᠰᠤᠭ	AMYAN-V USUG	HJ1	2	N	N+N	(DOTE JAM) GEDEG CAGAN TULOGAI-DV MINGGOL USUG SVR
агуу	ᠠᠭᠤᠭᠠ	AGVV	HJ2	1	A	A	ALTAI OBOG-VN HELEN E TURUG HELEN-U SINGELTE FRANSVW
агшиг	ᠠᠭᠰᠢᠭᠠ	AGSILG A	HJ1	1	N	N	ERTEN-U TURUG BICIO USUG-TU ER, E UGEN-DU ORODAG TVSHAI
амаа мэргэжлийн үг	ᠠᠮᠠᠵᠠᠪᠠᠷᠢᠭᠡᠯᠠᠨᠭᠡ	AME MERGEJL-UN UGE	HJ5	3	N	N+N+N	HOGOLAI-YIN HEI GAGCA HU AMA-BAR DAMEN GARHV ABASY, I
амаа нэрвйдгэл үг	ᠠᠯᠳᠠᠷᠨᠡᠷᠡᠶᠡᠳᠡᠬᠡᠯᠭᠡ	ALDAR NEREYDEH UGE	HJ3	3	N	N+N+N	ALDAR NEREYDEH UGE BOL, ULAN DOTOR, A-ACA ONCALAV TI
амаа үг	ᠠᠯᠳᠠᠷᠭᠡ	ALDAR UGE	HJ3	2	N	N+N	LSHY E NI BALDAN GEDEG NER, E BOL, HELE JU-YIN EGUDEN-EC
алтайн язгуур хэл	ᠠᠯᠲᠠᠭᠠᠨᠵᠠᠵᠠᠭᠤᠷᠬᠡᠯ	ALTAI-YIN JAGVR HELE	HJ6	3	N	N+N+N	TEGEGED BEJHU-DU BVI BOLHV DELBELUGCT AMY A NI AMHIV
ам	ᠠᠮ	AMA	HJ2	1	N	N	EGUN-DU HELE-YI AMAN DOTOR, A-BAN, HODELUL UGEI BAYTHV

Зураг 4. Монгол хэлний зүйн бүтээл дэх хэл шинжлэлийн нэр омьёоны салбар хөмрөг

Дүгнэлт

Хэлний материалын хөмрөг нь нэгэн том хэлбэрийн мэдлэгийн сан болсны хувьд, хэлний шинжлэлээс гадна нийгэм, амьдрал, улс төр, аж ахуй зэрэг олон талын асуудлыг судлахад мэдээлэл, тоо баримтаар хангаж өгдөг. Хөмрөгийг байгуулах зорилго бол хэлний бүтэц ба хэрэглээнд задлан шинжлэл хийж улмаар үндсэн судалгааны чанар чансаагаа сайжруулах явдал юм. Энэхүү өгүүлэлд өгүүлсэн “Монгол хэлний зүйн бүтээлийн хөмрөг”-ийг байгуулах бидний ерөнхий зорилго бол нэр томъёог компьютерт ялган таниулах (automatic term recognition) судалгаанд хэрэглэх гол хөмрөгийг бүрдүүлэх явдал боловч зөвхөн энэ зорилгоос өөр бусад олон төрлийн судалгаанд хөмрөгийг ашиглах боломжтой юм.

Б.Ринчен “Монгол бичгийн хэлний зүй” хэмээх дөрвөн боть бүтээл нь агуулгын хувьд авиа зүй, үсэг зүй, үг зүй, өгүүлбэр зүй зэрэг монгол хэл шинжлэлийн үндсэн ойлголтыг хамруулсан хэл зүйн бүтээл учир хэл шинжлэлийн нэр томъёо тус бүтээлд цөөнгүй байна. Эдгээр хэл шинжлэлийн нэр томъёоны салбар хөмрөгийг төгөлдөржүүлж электрон толь бичиг боловсруулах боломжтой.

Судалгааны үр дүнд Б.Ринчен “Монгол бичгийн хэлний зүй” дөрвөн боть бүтээлийн кирилл, монгол бичиг, латин галигаарх гурван хэлбэрийн “боловсруулаагүй хөмрөг” буй болов. Уг боловсруулаагүй хөмрөгт үг зүйн боловсруулалт хийж үг тус бүрд үгсийн аймгийн тэмдэг, үгийн хувиллын тэмдэг хадсан “боловсруулсан хөмрөг” буй болов. Цаашид энэхүү судалгаагаа үргэлжлүүлж “Монгол хэлний зүйн бүтээлийн хөмрөг”-т тулгуурласан нэр томъёог автоматаар таних дүрэм нөхцөл бичих, программ зохиох ажил; “Хэл шинжлэлийн нэр томъёоны салбар хөмрөг”-т тулгуурласан электрон толь бичиг зохиох ажлуудыг үе шаттайгаар хийж байна.

Ном зүй

Монгол хэлээр:

- Мөнхцэцэг, Н., Мөнх-Учрал, Э., Эрдэнэ-Очир, Т. (2022). *Монгол хэл судлал. XV боть. Хэрэглээний хэл шинжлэл, хэл боловсруулалт, түгээмэл хэл зүй*. Улаанбаатар: Соёмбо принтинг.
- Мөнх-Учрал, Э. (2020). Цахим орчин дахь Монгол хэл, бичгийн хэрэглээг медиа мониторингийн системээр шинжлэх нь. *Монгол судлал*, XLIV (543): 115-130.
- Насан-Урт, С. (2004). *Монгол хэл бичгийн сураг занги боловсруулах онол практикийн зарим асуудал*. Улаанбаатар.
- Ринчен, Б. (2015а), *Монгол хэлний зүйн бичиг. Удиртгал, Авиа, Үг, Өгүүлбэрийн зүй*. Улаанбаатар: Соёмбо принтинг.
- Ринчен, Б. (2015b). *Монгол бичгийн хэлний зүй. Удиртгал*. Улаанбаатар: Соёмбо принтинг.

Монгол бичгээрх ном, өгүүлэл:

- Dabqurbayar. (эмх.). (2018), *Тоуацаг-а келе шинжиле-үн удуридгал*. [Тооцоо хэл шинжлэлийн удиртгал] Kökeqota: Öbür mongγol-un yeke surγayuli-yin keblel-ün qoriy-a.
- Dabqurbayar. (2009), *Mongol kelen-ü ündüsün üyile kelkiče-yi compiyuter-tü taniyulqu sudulγan*. [Монгол хэлний үндсэн үйл хэлхийг компьютерт таниулах судалгаа] Kökeqota: Öbür mongγol-un arad-un keblel-ün qoriy-a.
- Rinčen, B. (2020). *Mongγol bičig-ün kelen-ü jüi. Uduridgal. Terigün debter*. [Монгол бичгийн хэлний зүй. Удиртгал. Тэргүүн дэвтэр] Kökeqota: Öbür mongγol-un arad-un keblel-ün qoriy-a.

- Wen Feng. (2020), *Mongyol kelen-ü ulus törü-yin ner-e tomiy_a-yi compiyuter-tü ilyan taniyulqu sudulyan*. [Монгол хэлний улс төрийн нэр томъёог компьютерт ялган таниулах судалгаа] Öbür mongyol-un yeke surayuli-yin magistr čola qamayalaqu ögülel.
- Xiao Juan. (2016), “*Nigen dabqur asar*”-un kelen-ü materiyyal-un bolbasurayulalta ba hereglege. [“Нэгэн давхар асар”-ын хэлний материалын боловсруулалт ба хэрэглээ] Öbür mongyol-un yeke surayuli-yin magistr čola qamayalaqu ögülel.

Гадаад хэлээр:

- Sandra Kubler & Heike Zinsmeister, (2015). *Corpus linguistics and Linguistically annotated Corpora*, London: Bloomsbury.

Abstract

The language text corpus is a large knowledge corpus and provides useful information and data for the study of many aspects of society, life, politics, and economy in addition to linguistics. The large collections of text in a language is collected to store in a computer to form a text corpus. A lot of rich language knowledge is hidden in the vocabulary created by this way, and in order to bring it out and put it into real use, the vocabulary needs to be developed. By the vocabulary, knowledge of real value for research and application can be extracted from language material. Therefore, researchers create various levels, and in order to find the information needed for their research, they make detailed dissections and markings in more than a level. By processing at many levels, such as lexicography, syntax, semantics, etc., it can be fully used in research such as dictionary creation, machine translation, automatic sentence generation, phonetic analysis, information search, and automatic recognition. In this article titled in «Automatic Recognition of Mongolian Linguistic Terms» would be considered about the compilation and processing of written language materials based on the four-volume book as «Grammar of Mongolian Writing Language» (Mongol bičgiin helnii züi) by Academician Byambyn Rinchen, and he created a branch of linguistic terms in that work.