**Оршил**

Шинжлэх ухаан, технологийн дэвшилттэй холбоотойгоор бичиг баримтыг цахим тооцоолуур ашиглан янз бүрийн хэрэгцээ шаардлагад нийцүүлэн боловсруулах болсноор мэдлэгт тулгуурласан нийгэмд, цахим хэл шинжлэлийн чиглэлийг хөгжүүлэхгүйгээр урагшлах боломжгүй болж байна. Иймээс олон улс орон өөрийн хэл, бичгийн цахим орчинд боловсруулах зүй тогтлыг нь цахим тооцоолуурт таниулахаар хүч хөдөлмөр, хөрөнгө зарцуулж, түүнийхээ хирээр дэлхийн хөгжилтэй хөл нийлэх боллоо.

Хүмүүний хэлийг цахим тооцоолуурт боловсруулснаар ямар ч улсын иргэд хоорондоо төрөлх хэлээрээ ойлголцох, техник технологийг шууд, хялбар ашиглах асар том боломжийг олгоно.

Бид бүгд одоо мэдээллийн хүчийг илүүтэйгээр ашиглан, түүний асар их урсгал дундаас мэдлэг бий болгож байна. Мэдээллийг илүү үр дүнтэй ашиглахын тулд эх хэлээрээ бүрэн дүүрэн орчуулж, ойлгох шаардлагатай. Эх хэлрүүгээ хөрвүүлэхдээ буруу, өөр утга санааг илэрхийлсэн үг ашиглах, эрлийз үг болгох явдал сүүлийн жилүүдийн бүтээлүүдэд гарч байгаа нь эх хэлний тогтолцоо алдагдах аюултай юм [1]. Тиймээс дэлхий дахинаа даяарчлал зарлаж буй цагт улс орны тусгаар тогтнолын нотолгоо болсон үндэсний хэл, бичиг, соёл түүх, өвөрмөц онцлогоо хадгалан үлдэх чухал гарц бол хэл, бичгээ мэдээллийн технологийн хурднаас хоцролгүй, шат ахиулан боловсруулах явдал юм.

Өндөр хөгжилтэй орнууд сүүлийн 20-иод жил энэ чиглэлээр ажилласны үр дүнд хэл сурах хялбар технологиос авахуулаад орчуулгын машин хүртэл бий болж эдгээр нь ач тусаа өгсөөр байна. Харин манай орны хувьд энэ чиглэл дөнгөж эхлэл үедээ байгаа бөгөөд монгол хэлийг машинаар боловсруулах багагүй судалгаа шинжилгээний ажил хийгдэж байгаа ба түүний нэгээхэн хэсэг нь монгол хэлний үгзүйн шинжилгээний чиглэл юм.

**Зорилго**

Дэлхийн олон хэлийг цахимын орчинд боловсруулж буй түгээмэл тогтолцооны хүрээнд монгол хэлийг мөнхүү арга зарчмаар боловсруулах асуудал чухал болж байна.

Эдгээр түгээмэл тогтолцооны нэг болох ПСи-Киммо програмд хэлний зөв бичих үгзүйн дүрмийг хоёр түвшинт загварт хөрвүүлэн дүрэм, үг, дагавар, нөхцөлийн сан (хэлний нөөцүүд) байгуулж өгснөөр үгийн шинжлүүрийг хийх боломжтой юм.

Энэ ажлын зорилго нь монгол хэлний үгзүйн дүрмийг аль болох хувирах үгнээс нь хамаарахгүйгээр зөв хувирах  үгзүйн хоёр түвшинт дүрмийг зохион туршиж, хэрэгжүүлэн, үр дүнг дүгнэх юм. Зорилгодоо хүрэхийн тулд дараах зорилтуудыг дэвшүүлэн тавилаа.

-          ПСи-Киммо програмын ерөнхий бүтцийг судлах

-          Хоёр түвшинт дүрмийг судлах

-          Монгол хэлний үгзүйн дүрмийг судлах

-          Бүтээврийг ангилах

-          Хоёр түвшинт дүрэм боловсруулах

-          Туршиж, дүгнэх

**Судлагдсан байдал**

Олон орны эрдэмтдийн боловсруулсан арга техник, хөгжүүлсэн програмуудаас Киммо Коскеннимигийн 1983 онд санаачилсан хоёр түвшинт үг зүйн дүрмийг хэрэгжүүлсэн ПСи-Киммо нь Финлянд[7], Турк[10], Англи[11], Япон гэх мэт олон хэлэнд туршигдсан үгийн шинжлүүр юм.

Харин Монгол хэлэнд ПСи-Киммо хоёр түвшинт үг зүйн талаар хэд хэдэн судалгаа хийж байсан бөгөөд анхных нь зөвхөн латин үсэг ашиглан үг зүйн дүрмийг загварчлах оролдлого байжээ [3]. Энэ нь цаашид ашиглахад бэрхшээлтэй байсан ба үүнийг крилл үсгэнд хөрвүүлэх шаардлагатай байв. Энд нийт 36 дүрэм боловсруулсан ба тэдгээр нь гээгдэх ба гээгдэхгүй эгшиг, эгшиг жийрэглэх ба эс жийрэглэх дүрмийг бүрэн гүйцэтгээгүй буюу бүх нөхцөлөөр хувиргаагүй, санамсаргүй цөөн хэдэн үгэнд туршсан дүрмийг сайн хөгжүүлэх хэрэгтэй байсан [15]. Үүний дараа ПСи-Киммо програмын үүсгүүр болон задлуурын хэсгийг Юникод дэмждэг болгон хөгжүүлснээр крилл үсгээр хоёр түвшинт дүрэм зохиох боломжтой болсон [4]. Ингэснээр монгол хэлээр хоёр түвшинт дүрмийг хэрэгжүүлсэн магистрын судалгааны ажил хийж эхэлсэн ба энэ нь зөвхөн нэрээс нэр үг үүсгэх дагавраар хувирах нийт 9 дүрэм боловсруулсан байсан [14]. Дээрх судалгааны ажлууд нь хувирах үгнээс хамааран нөхцөл бүрийг нэг бүрчлэн зааж өгсөн нь тухайн үгэнд ямар нөхцөл залгахыг мэдэж байх өгөгдлийг хэрэглэгчээс шаарддаг, эгшиг зохицох ёсыг хоёр түвшинт дүрмийн аргаар шийдэж чадаагүй гэх зэрэг дутагдалтай байсан юм.

Энэ нь хоёр түвшинт загварын боломжийг бүрэн ашиглаагүй юм. Иймд хоёр түвшинт дүрмийг монгол хэлэнд зохион турших хэрэгтэй болсон. Үүнээс гадна монгол хэлний үг зүйн хувиллыг толь бичигт суурилсан буюу “өгөгдөлд тулгуурласан тогтолцоо”-ны аргаар хэрэгжүүлсэн Ш.Чоймаагийн “Монгол хэлний хэлзүйн толь бичиг” -ийг машин орчуулгын түвшинд ашиглаж байгаа юм [12]. Энэ нь бүхэлдээ гар аргаар бий болдог тул тус салбарт ажиллаж буй хэлний мэргэжилтнээс их хэмжээний өгөгдөл доторх нэгж бүрийг бодож, няхуур хандахыг шаарддаг ба өгөгдөлд ороогүй үгийг хувиргах боломжгүй гэх мэт дутагдалтай.

**Судалгааны шинэлэг тал, ач холбогдол**

“Хиймэл оюун ухаан” бүхий цахимд зориулсан судалгаа нь “хүмүүний хэл” (natural language)-ний зүй тогтол, үйлдэл бүрийг нарийн тодорхойлох шаардлагатай байдаг тул уламжлалт судалгааны онолыг үг болон хэлзүйн хэлбэр бүр дээр хэрхэн бууж байгааг тооцоолж, түүнийгээ загварчлах нөхцөл болзол тулгадаг. Тэрхүү нөхцөлийг хангасан өгөгдөл бэлэн бүрдсэн тохиолдолд цахим тооцоолуур зөв ажиллана. Эл чиглэлээр урьд хийгдсэн ажлын үр дүнгээс үзэхэд монгол хэлний уламжлалт аргаар тодорхойлсон дүрмийг хүмүүний хэлний боловсруулалтын технологид шууд тавих нь нэн төвөгтэй байгаа юм. Тиймээс уг ажлаар монгол хэлний үгзүйн дүрмийг улам илүү нарийвчлан судалж, хувирах үгнээс хамаарахгүйгээр нэг төрлийн нөхцөлийг ерөнхийлөн нэг тэмдэглэж оролтод өгснөөр бидний зохиосон хоёр түвшинт дүрэм зөв хувиргах, хоёр түвшинт загварын бүх боломжийг ашиглан өмнөх судалгааны дутагдлыг засаж, сайжруулахын зэрэгцээ эгшиг зохицох ёсыг дүрмийн аргаар боловсруулан гүйцэтгэх нь толь бичигт суурилсан монгол хэлний үг зүйн хувиллын нүсэр ажлыг хялбар, хурдан загварчилсан шинэлэг ажил болно. Ерөнхийдөө цахимд зориулсан хэлний тооцоолол, өгөгдлийг эхний шатанд бий болгоно. Дараагийн шатанд эхний боловсруулалтад тулгуурлан эргээд хүнд чиглэсэн хэрэглээг бий болгодог.

Монгол хэлний үг зүйд хоёр түвшинт дүрмийг бүрэн боловсруулснаар олон талын практик ач холбогдолтой. Үүний заримаас дурдвал:

* Зөв бичгийн шалгуур програм бүтээх
* Машинаар үгийн сан үүсгэх
* Машин орчуулгын түвшинд ашиглах
* Монгол хэлний сургалтанд ашиглах

**НЭГДҮГЭЭР БҮЛЭГ ХЭЛ БОЛОВСРУУЛАЛТЫН ОНОЛЫН ОЙЛГОЛТ**

Компьютер, мэдээллийн технологийн салбар нь бусад шинжлэх ухаан, нийгмийн салбарын хэрэгцээнд байнга ашиглагдаж, бүх шинжлэх ухааны салбаруудтай нэгдэж уулзвар салбарыг үүсгээд байгаа билээ. Сүүлийн үед эрчимтэй хөгжиж байгаа эдгээр салбарын нэг бол цахим хэл шинжлэл юм.

* 1. **Цахим хэл шинжлэлийн тухай**

Харьцангуй шинэд тооцогдох энэхүү салбар шинжлэх ухааныг орчин үед дараах байдлаар тодорхойлж байна. Үүнд:

“Цахим хэл шинжлэл гэдэг бол тооцооллын үүднээс хүмүүний хэлийг ойлгох, боловсруулах, бүтээж гаргах шинжлэх ухааны судалгаа бөгөөд хэл шинжлэлийн янз бүрийн онцгой үзэгдлийг тооцооллын загварт оруулах зорилготой, олон шинжлэх ухааны уулзвар дээр байдаг, хиймэл оюун ухааны салбар болсон ухаан юм” [12].

Энгийнээр бол компьютер (цахим тооцоолуур) –ын тусламжтайгаар хүний хэлийг судалж, түүний зүй тогтлыг оновчтой тодорхойлон хэлний хэрэглэгдэхүүн бий болгож улмаар машинд хүний хэлийг суулгах зорилготой юм. Машин хүний хэлийг эзэмших нь хүн төрөлхтөн технологийг хамгийн үр дүнтэй ашиглах нэг том алхам юм.

Цахим хэл шинжлэлийг нэгд, хэл шинжлэлд зориулж цахим тооцоолуураар боловсруулсан арга барил, түүний хэрэглээ хоёрт, цахим тооцоолуурт зориулж хэл шинжлэлийн талаас бүтээсэн арга барил, түүний хэрэглээ гэсэн хоёр талаас нь авч үзэж болно [12].

* Эхнийх нь хүмүүс хэлээ мэдэж байхын тулд юу мэдэх хэрэгтэй, түүнийгээ хэрхэн ашиглаж, юу хийж бүтээх вэ гэдэгт чиглэнэ.
* Дараах нь цахим тооцоолуурыг яриа ойлгодог, хэлийг хооронд нь орчуулдаг болгох зэрэгт гол анхаарлаа хандуулдаг. Энэ хүрээнд хамаарах судалгааг цахимд зориулж хэлийг судлах, цахим тооцоолуурыг хэлний хувьд хүнтэй адил түвшинд харилцдаг болгохын тулд хэлний тогтолцоог бүх түвшинд нь бүтээхэд чиглэсэн судалгаа юм. Хэлний тогтолцооны түвшин дараах бүтэцтэй байна. Зураг 1-д үзүүллээ.

***Зураг 1****. Хэл боловсруулалтын тогтолцооны түвшин*

Хэл боловсруулалтын тогтолцооны түвшин бол бодит байдалтай холбоотойгоор тухайн хэлийг боловсруулахын тулд энгийнээс эхлээд ярианы хэллэг хүртэл түвшин бүрийг авч үзэн түүний мөн чанарыг мэдэж, тухайн хэлний учрыг олох явдал юм.

Цахим хэл шинжлэл хэрэглээнээсээ шалтгаалан аль нэг дэд түвшин, тэдгээрийн хослол ашиглан, янз бүрээр авч судалдаг. Жишээ нь үг зүйн түвшинг үндэслэн тухайн хэлний үгийн язгуур үндсийг олох, үг хувилгах гэх мэт програмд ашигладаг. Түвшингүүдийн хоорондын ялгааг зөв гаргаж, нарийн тодорхой холбосноос хэл боловсруулалтын чанарыг үнэлдэг.

Цахим хэл шинжлэлийн хоёр гол хандлага нь *“дүрэмд суурилсан тогтолцоо”* (Rule-Based Systems), *“өгөгдөлд тулгуурласан тогтолцоо”* (Data-Driven Systems) юм [12].

*“Өгөгдөлд тулгуурласан тогтолцоо”* нь хэл шинжлэлийн мэдлэгийг далд илэрхийлсэн кодчилолтой, статистик буюу цахим тооцоолуур ашиглах аргад тулгуурласан ч өгөгдлийн үлэмж сурвалж шаардаж, түүнийг бүрэн, зөв оруулахад хүний ажиллагаа маш их тулгарч, үр дүн нь зөвхөн өгөгдлийн сангийн цар хүрээнээс шууд хамааралтай.

*“Дүрэмд суурилсан тогтолцоо”* нь хэл шинжлэлийн мэдлэгийг тодорхой илэрхийлсэн кодчилолтой, тухайн хэлний ерөнхий зүй тогтолд нийцүүлэн боломжит хувилбаруудаар шалгагдан, гараар оруулсан хэлзүйн дүрэмд тулгуурлана. Ингэснээр турших болон алдаа илрүүлэхэд хялбар байдаг. Хүний ажиглах, нэгтгэн дүгнэх их хичээл зүтгэлээр бүтээнэ. Оруулж буй өгөгдлийн цар хүрээ өргөсөх тусам бэрхшээлтэй болж ирэхийн дээр, хэрэвзээ оруулсан дүрэм бүх нөхцөлийг хангаж чадсан бол ямар ч өгөгдөл буруу гарахгүй, хэрэглээний найдвартай, хүртээмжтэй байдлыг хангаж чадах боломжтой юм.

* + 1. **Цахим хэл шинжлэлийн хөгжил ба түүний салбарууд**

Хэлийг боловсруулах, хэлний хэрэглээг хангахад цахим тооцоолуур ашиглах санаа төрсөн шалтгаан нь хүний мэдлэгийг ашиглахад цаг, хөлс, оюуны болоод төлөвлөгөөт ажлын боломжоос хараат буюу товчоор, их үнэтэй байдаг бол цахим тооцоолуур арифметик үйлдлийг хүнээс илүү хурдан, оновчтой гүйцэтгэж чадах нь нотлогдсоор илүү бага зардал зарцуулж болох юм гэж үзсэнд оршино. Үүний хамгийн их нухацтай зүйл бол хэлний мөн чанар юм. Хүний хувьд хэдий энгийн санагдаж болох ч энэхүү үйдлийн бүтэц бүрэлдхүүнийг цахимд ойлгуулахад тийм хялбар зүйл биш ажээ. Хэлийг цахимд боловсруулах зорилтыг биелүүлэхийн тулд хэлний уул язгуураасаа нарийн хүний ухамсар шаардсан шинжтэй холбогддог учир энэ судалгаа нь 1940 оноос эхлэн эдүгээг хүртэл олон жил болсноор зогсохгүй, зорьсон төгс үр дүндээ бүрмөсөн хүрээгүй байсаар байна.

Одоогоор яриа таних ба үүсгэх зөв бичих зүйн болон хэлзүйн алдаа засах, цахим орчуулга, мэдээлэл хайх, сэрээх зэрэг хэл боловсруулалтын хэрэгцээг хангах програм хангамжийн болон техник хангамжийн дэвшлүүд гарсаар байна.

Цахим хэл шинжлэлийн туйлын зорилго нь машинаар хүн шиг хэл боловсруулахад оршино.

**Цахим хэл шинжлэлийн салбарууд**

Цахим хэл шинжлэлийг судлагдахуун нь “яриа” эсвэл “бичвэр” болж байгаагаас хамааран нэг нь бичих, ярих үүрэгтэй ижил бол нөгөө хэсэг нь унших, сонсох үүрэгтэй ижил юм. Иймээс дараах хоёр үндсэн салбарт хуваан үзэж болох юм.

* Яриа танихуй (Speech recognition) буюу яриа бий болгох нь ярианы хэлийг ойлгох, ярианы хэлнээс бичвэр үүсгэхтэй холбоотой салбарууд.
* Хэлзүйн задлал хийх (Parsing) ба хэлзүйн дагуу бүтээх (Generation) нь хэлийг тусгаар хэсгүүдэд задлах буюу нэгтгэхтэй холбоотой салбарууд.

**ЦХШ-ын дэд салбарууд:**

* Хүмүүний хэлний боловсруулалт (Natural Language Processing)
* Тооцоолын утгазүй (Computational Semantics)
* Хөмрөгийн хэл шинжлэл (Corpus Linguistics)
* Хэлзүйн задлуурын зохиомж (Design of Parsers)
* Тэмдэглүүрийн зохиомж (Design of Taggers)
* Цахим орчуулга (Machine Translation)

Дээрх хоёр үндсэн салбар нь хэлийг шинжлэх үүрэгт төвлөрсөн ба түүний хэрэгцээ нь:

* Бичгийн, ярианы хэлэнд дүрмийн алдаа, учир дутагдал байгааг олох
* Нийгэмд хэл бичгийн соёл хэрхэн өөрчлөгдөж байгааг харах
* Энгийн програм хангамжид ашиглах (зөв бичих дүрмийн алдаа засах, ярианы, орчуулгын програм, толь бичгийн судалгаа)
* Шинжлэх ухаан үйлдвэрлэлд нэвтрүүлэх (хамгаалалт, удирдлагын систем, автомашины электроник хиймэл оюун, робот)
* Үндэсний аюулгүй байдалд ашиглах (сансар судлал, цаг агаар зэвсэгт хүчин) боломжийг олгоно.

Хэлний технологи ерөнхийдөө бол яриа ойлгож эхлээд яриа үүсгэж дуусдаг. Эдгээр бүх л түвшингүүд их бага хэмжээгээр сэтгэл судлалын салбарыг ашиглан тооцох шаардлагатай байдаг. Гэсэн хэдий ч хэл шинжлэлд сэтгэл судлалыг онцгойлон анхааралдаа авч тооцож чаддаггүй байна.

* + 1. **Цахим хэл шинжлэлд шаардагдах мэдлэг**

Хэл боловсруулалтын гол цөм нь эх хэлний бүхий л түвшний хангалттай мэдлэгийг судлан, хоорондын уялдааг гаргах шаардлагатай. Хэл боловсруулт нь дараах түвшингүүдтэй. Үүнд:

***Авиа зүй*** (phonology )

Авиа хэллэгийн боловсруулалт хийх буюу хэл яриаг бүтээж байгаа авиаг тал бүрээс нь судалдаг хэл шинжлэлийн нэг салбар ухааныг авиа зүй гэнэ. Дан ганц энэ түвшингээс бүрэн цогц програм үүсэх тохиолдол бараг үгүй. Энэ салбарын гол судалгаа нь үгийн хязгаарыг олж, таних явдал юм.

***Үг зүй*** (Morphology)

Үгийн үе, залгавар, угтварын боловсруулалтыг хийх буюу үг бүтээх дагавар, нөхцөлийг тусгайлан судалж цахим хэл шинжлэлийн дүрмийг боловсруулдаг. Монгол хэл нь Алтай язгуурын залгамал хэл бөгөөд маш олон дагавар, нөхцөлийг ар араас залгаж үг бүтээдэг. Энэ түвшинд зөв бичих үсгийн дүрмийг түлхүү хэрэглэдэг. Хамгийн гол нь үг зүйн задлан ялгал, үгийг боломжит хувилбараар зөв хувилдаг байх хэрэгтэй.

***Өгүүлбэрзүй*** (Syntax)

Үг хэллэгийн хоорондын бүтцийн харилцааг олж, өгүүлбэрийн бүтэц, дүрмийг зохион байгуулах үйл ажиллагаа, боловсруулалтыг хийх

***Утгазүй*** (Semantic)

Өгүүлбэр болон үгийн утгын боловсруулалт буюу хэд хэдэн өгүүлбэрийн хооронд ажиллаж анализ хийдэг. Тухайн үг ямар утгаар орж байгааг, өмнөх өгүүлбэрт орсон үйлийн эзэн мөн биш, тухайн нэр үг хийсвэр зүйлийг илтгэж байгаа эсэх, үнэхээр бодитой юмыг зааж байна уу, үгүй юу гэхчлэн маш олон талаас нь утгыг бодож олно.

***Найруулгазүй*** (Stylistics)

Тухайн нэг сэдвийн зохицол буюу үг илүүдсэн эсвэл дутсан, хэлзүйн дагуу зөв хэлбэржсэн эсэх, найруулгын түвшний уялдаа зэрэгтэй холбоотой мэдлэг нь бусад түвшний мэдлэгийн хамтаар хэрэгжих боломжтой ахисан түвшний боловсруулалт юм.

Тэгэхээр цахим хэл шинжлэлд эдгээр мэдлэгүүдийн иж бүрэн өгөгдөл шаарддаг тул харилцан уялдсан арга барил өгөгдлийг бүтээх нь чухал билээ.

* 1. **Үгзүй (Morphology)**

Оюун ухаант хүний хөгжлийн амин сүнс болсон хэлний гол цөм нь яах аргагүй үг зүй юм. “ Хэлний ертөнцийн төв нь үгзүй бөгөөд хэлийг танин мэдэхийн тулд юуны өмнө түүний үгзүйн байгууллыг судлах хэрэгтэй ”[3] гэж тэмдэглэсэн байдаг.

Үгзүй нь дээр өгүүлсэнчлэн үгийн бүтэц ба үгийн хэрхэн хэлбэржихийг судлах цахим хэл шинжлэлийн салбар юм. Үгзүйн ойлголтыг гаргахын тулд дараах үгнээс үгзүйн шинж чанарыг нь ажиглавал:

* “морийг” гэсэн энэ үг “морь” гэсэн үгийн хувирсан хэлбэр
* “морийг” бол “морь” гэсэн үгийн үндэс дээр заахын тийн ялгалын “ийг” нөхцөл нэмэгдэж хувирсан байна.
* Үгийн үндэс “морь” гэсэн үгийн сүүлийн ь - и болж хувирсан байна.

Дээрх гурван зүйлийн хоорондын нарийн хэлхээ холбоосыг буюу нэгдүгээрт яагаад “ийг” гэсэн нөхцөлийг авч байгаа, хоёрдугаарт тухайн үгний гаралт ямар үед зөв болох, эцэст нь үгзүйн дүрмийг хэрхэн зохиох гэсэн асуултууд тулгарч байна. Тиймээс

үгзүйд дараах зүйлүүдийг авч үзэх нь харагдаж байна.

1. Хамгийн бага утгат хэсгээс хэрхэн үг бүтээх
2. Үгзүйн дүрмийг боловсруулах
3. Утгат хэсэгт залгах залгаврын ялгаа, дарааллыг тодорхойлох

**1.2.1 Үгзүйн процесс**

Үгзүйн процесс нь хамгийн бага утгат хэсэг тухайн хэлний үгзүйн дүрмийн дагуу хэлбэржих явц юм. Утгат хэсэгт язгуур ба залгавар бүтээвэр хоёр хамаарна. Язгуур ба залгавар бүтээвэр хоёрын хоорондын үндсэн ялгаа нь залгавар бүтээвэр язгуурт нэгдэх ба хувирсан хэлбэр нь үгийн үндэс болж болох ч дахин залгавар авахдаа хязгаартай байдаг. Харин язгуур нь залгавар авахдаа хязгаарлалт байхгүй. Мөн үгзүйн процесс хэлний хэв шинжээс ихээхэн хамаарна. Хэлний хэв шинж ерөнхийдөө 3 төрөл байдаг. Нэгд, утгат хэсгийг хэлхсэн (morphemes concatenative) хэл. Жишээ нь англи хэл “**antidisestablishmentarianism”.** Хоёрт, залгамал хэл буюу бүтээврийг боломжит хэлбэрээр залгасан урт үгзүйн процесс үүсэх хэл юм. Англи хэл шиг угтвар (prefix) байхгүй гэсэн үг юм. Жишээ нь Турк, Монгол хэл “**хадгалагдчихсан”.** Гуравт, загвар хэл буюу утгат хэсгүүд нь салангид, загварт үгзүйн процесс үүсэх хэл. Жишээ нь Араб хэл.

Үгзүйн процесс нь төрөл бүрийн боловсруулалтыг шаарддаг. Үүнд:

* *Оруулж тэмдэглэх (infixation):* Үгийн утгат хэсэг дээр залгавар залгахад өөр ямар нэгэн үсэг нэмэгдэж орох үед
* *Тойрч тэмдэглэх (circumfixation):* Утгат хэсгийг бүтэн тойроод тасалдаж өөр үсэг хасагдаж, нэмэгдэхэд
* *Давталт (reduplication):* Утгат хэсгийн тухайн нэг үе дахин давтагдаж нэмэгдэхэд
* *Давхарлах (gemination):* Дээрх боловсруулалттай төстэй ба зөвхөн үсэг давхарлаж ороход
* *Чөлөөтэй тархсан (degemination):* Давталтаар нэмэгдсэн үетэй хамааралтай дахин үе нэмэгдэх
* *Солилцох (metathesis):* Хоёр үсгийг хооронд нь эсвэл өөр үсгээр солиход
* *Устгах (deletion):* Утгат хэсэг дээр залгавар залгахад үсэг устах зэрэгт үйлдлүүдийг хэрэглэдэг.

Энэхүү цахим үгзүй хөгжил 1960 оноос эхэлсэн ба олон орны эрдэмтдийн боловсруулсан арга техник, хөгжүүлсэн програмуудаас Киммо Коскеннимигийн 1983 онд санаачилсан хоёр түвшинт үгзүйн дүрмийг хэрэгжүүлсэн ПСи-Киммо үгийн шинжлүүр олон хэлэнд туршигдсан юм. Хоёр түвшинт үгзүйн загварыг бусад цахим үгзүйн загваруудтай харьцуулсан тухай дараагийн бүлэгт дэлгэрэнгүй тайлбарлана.

**ХОЁРДУГААР БҮЛЭГ ХОЁР ТҮВШИНТ ҮГ ЗҮЙН ШИНЖЛҮҮР**

Хоёр түвшинт үгзүйн шинжлүүр [9] бол цахим хэл шинжлэлээр хөгжүүлэгдэх хэл шинжлэлийн хэрэгсэл юм. Үүнийг цаашлаад хүмүүний хэл боловсруулалтын системд хэрэглэдэг. Үгзүйн шинжлүүр нь үгийг хэл зүйн зөв бичих дүрмийн хувьд задлан, хэл зүйн хамгийн бага утгат хэсгээр харуулдаг, утгат хэсгүүдэд бүтээврийг залган үүсгэдэг програм юм. ПСи-Киммо програмд хэлний зөв бичих үсгийн дүрмийг хоёр түвшинт загварт хөрвүүлсэн дүрэм, үг, дагавар, нөхцөлийн сан (хэлний нөөцүүд) байгуулж өгснөөр үгийн шинжлүүрийг хийх боломжтой юм. Коскеннимийн хоёр түвшинт загварыг хөгжүүлсэн дэвшилттэй тал нь үнэхээр практик сайтай ба өндөр түвшний залгамал хэлийг үгзүйн дүрмийн дагуу сайн хувиргадаг.

Коскеннимийн энэхүү загвар нь хэв шинжит илэрхийллээр хөгжүүлэгдэх цахим хэл шинжлэл ба үгзүйн дүрэм хоорондын уламжлалт онцлог шинжид үндэслэгдсэн. *Хэв шинжит илэрхийлэл* нь бичвэрийг боловсруулах, хайлт хийх зэрэг компьютерын салбарт өргөн ашиглагддаг нэгэн хэл юм. Энэ ч утгаараа хэв шинжит илэрхийлэл хүмүүний хэл боловсруулалтын үндсэн технологийн хэрэглээнүүдийн нэг байдаг.

Аливаа хэл нь өмнөх цогцолборт дурдсан шиг энгийн байхгүй, харин ч илүү төвөгтэй, нарийн байдаг. Хэв шинжит илэрхийлэл нь тэмдэгтийн цуваагаар илэрхийлэгдэх төвөгтэй хэлийг бүрэн тодорхойлохын тулд нийтлэг тусгай тэмдэгт, операторуудыг ашигладаг.

Коскеннимийн хоёр түвшинт загвар хоёр бүрэлдэхүүн хэсгээс хамаардаг.

* Дүрмийн хэсэг, үг зүйн дүрмийг хэв шинжит илэрхийллээр илэрхийлсэн, төгсгөлөг төлөвт төхөөрөмж
* Үгийн сангийн хэсэг, үгийн утгат хэсэг ба үгсийн, залгавар бүтээврийн жагсаалт, түүнд үндэслэсэн хэлбэрүүд тэдгээрийг кодоор шифрлэх

Хоёр бүрэлдхүүний хамтын ажиллагаа нь үг үүсгэх (generation), үгийн задлуур (recognition) хоёрыг гүйцэтгэдэг. Хоёр түвшинт дүрмийг ашиглаж, шинжлүүрийн үгийн сангийн файлын үндсэн хэлбэрийг тодорхойлсны дараа тухайн хэлний *хоёр түвшинт үгзүйн шинжлүүр* болно.

Энэ загварын эхний түвшин нь үгийг шууд буюу үгийн сангийн утгаар илэрхийлсэн бол дараах нь үгзүйн дүрмийн дагуу зөв хувирах хэлбэрийг илэрхийлнэ. Товчоор хэлбэл үндсэн ба хувирсан хэлбэрүүдийн хоорондын харилцан холбоосын ***Хоёр түвшинт илэрхийлэл*** юм. Жишээ нь “түмний” үгийг хоёр түвшинтээр илэрхийлбэл:

Үгийн сангийн хэлбэр (Lexical form): т ү м э н + Ы н

Хувирсан хэлбэр (Surface form): т ү м 0 н 0 и й болно.

Энэ програм нь үгийн хувирсан хэлбэрийг авч үгийн үндэс, дагавар, нөхцөлөөр нь задалдаг задлуурын хэсэг, үгийн үндэс, дагавар, нөхцөлийг нь салгаж өгөхөд нийлүүлж зөв хувирсан үг гаргах үүсгүүрийн хэсэг, хоёр түвшинт дүрэм, үгийн санг уншиж санах ойд зохион байгуулах хэсгээс бүрддэг. (Зураг 2)

**Зураг 2.** ПСи-Киммо системийн үндсэн бүтэц

* 1. **Хоёр түвшинт дүрэм**

Энэхүү системийн дүрмийн хэсэг нь тухайн хэлний үгзүйн дүрмийг нарийвчлан гаргаж, хоёр түвшинт дүрмийн тэмдэглэгээг ашиглан шугаман хэлбэрт оруулсан файл болно. ПСи-Киммо, хоёр түвшинт дүрмийг сонгох нь давуу тал, практик ач холбогдол ихтэй юм.

**2.1.1 Хоёр түвшинт дүрмийн давуу тал**

Тэгвэл хоёр түвшинтээс өөр үүсгүүрийн ерөнхий дүрмийн хэлбэр, шинжийг авч үзье. Үүсгүүрийн дүрэм бол дараалан жагсаасан дахин боловсруулах дүрмүүд болно. Үүнийг дараах байдлаар тайлбарладаг.

1. Дахин боловсруулах дүрмүүд гэдэг нь ямар нэгэн тэмдэгтийг өөрчлөх эсвэл хувиргах дүрмүүд юм. Жишээ нь a>b уруу өөрчлөлтийн дахин боловсруулах дүрэм нь a, b тэмдэгтийн хоорондын холбоосоор a тэмдэгтийг дахин боловсруулах эсвэл эргүүлэх аргаар динамикаар өөрчлөх дүрмийг зохиодог. Энэ дүрмийн үйлдэлээ хийсний дараа тэмдэгт a алга болж өөр дүрэм хэрэгжүүлэх боломжтой болно.
2. Хэл шинжлэлийн онолын үүсгүүрийн дүрэм бол шат дараалсан процессоор явагдаж, *дараалсан* (sequentially) байдлаар хэрэглэдэг. Иймд дүрэм бүр дээр гаралтын илэрхийллийн шинэ дунд түвшин үүсдэг. Энэ дунд түвшинд эхний оролттой адил дараагийн дүрэм үйлчилдэг.
3. Үүсгүүр шинжлүүрийн дүрмүүд эрэмбэлэгдсэн: дүрмийн хэрэглэгдэх дарааллыг төрөл зүйлээр нь эрэмбэлдэг.

Эдгээрээс харахад үүсгүүрийн дүрэм нэг дүрэм хэрэгжүүлсний дараа өмнөх байдал нь хаягдаж дахин хандах боломжгүй болох, дүрмийг зөв дарааллаар эрэмбэлэхгүй бол худал үр дүн гарах дутагдалтай талуудтай. Харин хоёр түвшинт дүрэм эдгээр дутагдлыг зассан илүү сайн боловсрогдсон хэв шинжтэй дүрэм гэдгийг дараах байдлаар илэрхийлдэг.

1. Үүсгүүрийн дүрэм нь дараалсан эрэмбээр хэрэглэдэг бол хоёр түвшинт дүрмүүд нэгэн зэрэг хэрэглэдэг ба зэрэгцээгээр үйлчлэн зөв үр дүнг гаргах боломжтой.
2. Дараалан хэрэглэсэн үүсгүүрийн дүрэм нь үүсмэл, дунд түвшингүүд үүсгэдэг бол, хоёр түвшинт дүрмүүд зөвхөн илэрхийллийн хоёр түвшинг шаарддаг: үндсэн буюу үгийн сангийн, хувирсан гэсэн хоёр түвшин. Тэдгээр нь үүсмэл дунд түвшин биш. Ийм утгат загварыг хоёр түвшинт гэнэ.
3. Үүсгүүрийн дүрмүүд үндсэн тэмдэгтийг олон хувиргаж, дунд түвшин бүр дээр хувирсан тэмдэгтийг дахин боловсруулах дүрэм шаарддаг бол харин хоёр түвшинт дүрмүүд шууд баримтын үндсээр үндсэн ба хувирсан түвшингүүдийн хоорондын холбоосыг тодорхой илэрхийлж, түүнд тохирох статик харилцан холбоотой нэг л дүрэм шаардана.
4. Үүсгүүрийн дүрмүүд зөвхөн үүсмэл үе бүрийн дундаж хэлбэрт ханддаг бол хоёр түвшинт дүрмүүд үндсэн ба хувирсан хоёр түвшингийн орчинд л ханддаг. Үүсгүүрийн дүрмүүд үндсэн орчинг эргэж харах эсвэл хувирсан орчинг урьдчилан харж чадахгүй. Тодруулбал, хоёр түвшинт дүрмийн орчин нь хувирсан түвшин үгсийн сантай харилцан холбоотой илэрхийлэгддэг.
5. Үүсгүүрийн дахин боловсруулах дүрмүүд бол нэг чиглэлийн (үндсэн түвшингээс хувирсан уруу чиглэсэн), хоёр түвшинт дүрмүүд бол хоёр чиглэлийн шинжтэй. Хоёр түвшинт дүрмүүд үндсэнгээс хувирсан чиглэл (үүсгүүр горим) ба хувирсангаас үндсэн чиглэл (задлуур горим) болно.

Иймд үүсгүүр горимд хоёр түвшинт дүрмүүд оролтоор үндсэн хэлбэр ба гаралт нь хувирсан хэлбэр болоход задлуур горим нь үүний эсрэгээр нь хэрэгждэг. Хоёр чиглэлт шинжлүүрийн дүрмүүдийн туршилт нь маш тодорхой: хоёр чиглэлтийн дүрмийн тооцооллын гүйцэтгэл нь үг бүтээх үүсгүүрийн горимд хязгаарлагдахгүй, үгсийг задлан ялгалал хийхэд задлуурын чиглэлд бас ашиглана.

**2.1.2 Хоёр түвшинт дүрмийн тэмдэглэгээ**

Хоёр түвшинт дүрмийн хийсвэр тэмдэглэгээнүүд нь гурван хэсгээс бүрддэг. Үүнд: харилцан холбоосын боловсруулалтын, боловсруулалт хийгдэх газар буюу орчны тодорхойлолт, дээрх хоёрыг холбосон хамаарлын оператор болно.

*Харилцан холбоосын боловсруулалт*: Хоёр түвшинт дүмүүд үг бүрийг үгийн сангийн хэлбэр (ҮХ) ба хувирсан хэлбэр (ХХ) хоорондын харилцан холбоосоор авч үздэг. Жишээ нь үгийн сангийн хэлбэр багш+Ын ба хувирсан хэлбэр багшийн -ыг хэлэлцье.

ҮХ: б а г ш + Ы н

ХХ: б а г ш и й н

Үгийн сангийн ба хувирсан хэлбэрийн тэмдэгтийн хос бүр харилцан холбоосын хослол болох ба энгийн, өвөрмөц харилцан холбоосууд болно. Харилцан холбоосыг **үгийн сангийн тэмдэгт : хувирсан хэлбэрийн тэмдэгт** гэж тэмдэглэн бичдэг. б:б, a:a, Ы:й, +:и гэх мэт.

Харилцан холбоос нь яг тодорхой нэгээс нэгд харьцуулагдсан байх ёстой. Харилцан холбоос нь хоёр хэлбэртэй. Эхнийх нь б:б, a:a нь энгийн харилцан холбоос дараах нь Ы:й, +:и нь өвөрмөц харилцан холбоос болно. Энгийн ба өвөрмөц харилцан холбоосуудын нийлбэр нь тодорхойлсон боломжит хослолуудын олонлогоос бүрдэнэ. Өөрөөр хэлбэл бүх боломжит хослолууд энэхүү хоёр харилцан холбоосуудын аль нэгээр нь тодорхойлогдон зарлагдсан байх ёстой.

*Боловсруулалтын орчинг тодорхойлох:* Орчин нь боловсруулалт явагдах байрлалыг заах бөгөөд зүүн ба баруун гарын орчны тодорхойлолтыг сонгож болдог. Ядаж нэг нь заавал байх хэрэгтэй ба хоёр талын хоёуланг нь тодорхойлж ч болно. Орчин нь бас хоёр түвшинт харилцан холбоосын нэр томъёогоор дүрслэн тодорхойлно. Дүрмийн орчин (нөхцөл) нь тухайн хэлний цагаан толгойд багтсан үсгүүд, ямар нэгэн тэмдэгтүүд (@, #) ашиглан бичдэг. Ямар нэгэн тэмдэгт нь дүрмийн харилцан холбоосын хэсэгт бус дүрмийн нөхцөлийн хэсэгт л зөвхөн ашиглана. Жишээ нь:

Дүрэм 1. э эгшиг гээгдэх дүрэм

RULE э:0 => [V [MN|ZZ]|MN ZZ] \_\_\_ [MN|ZZ] +:@ [в#|Ү:@|Ы:@|У:@]

Дүрэмд доогуур зураас нь дүрэм хэрэгжих байрлал ба доогуур зураасны баруун, зүүн талыг дүрмийн орчин гэж нэрлэж байна.

*Хамаарлын оператор тодорхойлох:*  Дүрэм 1-ийн энэ = > хэсэг бол хамаарлын оператор буюу дүрмийн үйлдэл болно. Тэгэхээр энэ үйлдэл бол сум шиг хэлбэртэй түүний утга нь үүсгүүрийн дүрмийн дахин бичих утгатай сумнаас маш их ялгаатай юм. Дүрмийн үйлдэл нь илэрхийлж байгаа орчин ба харилцан холбоосын хоорондын харьцаагаар тодорхойлдог. Дүрэм нь дараах 4 үйлдэлтэй: =>, <=, <=>, \<=. Дүрмийн үйлдлүүдийн утгыг товчоор хэлбэл дараах утгатай:

=> тухайн орчин дахь харилцан холбоосд ганцхан үйлдэгддэг.

<= тухайн орчин дахь харилцан холбоос бүгдэд үйлдэгддэг.

<=> тухайн орчин дахь харилцан холбоосын дээрх хоёр үйлдэл хоёулангийн нь давхацсан нөхцөлд үйлдэгддэг.

\<= тухайн орчин дахь харилцан холбоос үйлдэгдэхгүй.

Хоёр түвшинт дүрмийн чухал шинж чанар бол үгийн сангийн хэлбэрийн тэмдэгт хувирсан хэлбэрийн тэмдэгтийн хоорондын харилцан холбоо **нэгээс нэгрүү** байхыг шаарддаг. Үгийн сангийн ба хувирсан хэлбэр хоёулангийнх нь тэмдэгтийн тоо тэнцүү байх ёстой ба үгийн сангийн тэмдэгт бүр хувирсан тэмдэгтэй яг нарийн харьцуулагдах ёстой. Шинжлүүрийн процесс буюу устгах ба оруулах үйлдэл нь тэг ба хоосон (NULL) тэмдэгтэй харилцан холбоосоор хоёр түвшинт загварт илэрхийлэгдэнэ. Үүний дагуу үгийн сангийн хэлбэрийн + (утгат хэсгийн зааг) нь хувирсан 0 тэгэд хамааралтай (энэ бол устгах үйлдэлд), хувирсан тэмдэгт болох ямар нэгэн үсэг нь (э) үгийн сангийн 0 тэгэд хамааралтай (энэ бол оруулах үйлдэлд). Жишээ нь оруулах үйлдлийг илэрхийлбэл:

ҮХ: с а н 0 + Ы н

ХХ: с а н г и й н

Хоосон (NULL) тэмдэгтийг бол зөвхөн дүрмийн дотоодод ашиглана. Гаралтын хэлбэрт хэвлэгдэхгүй ба оролтын хэлбэрт бичих шаардлагагүй. ПСи-Киммо нь үгийн сангийн оролтын хэлбэр с а н + Ы н ба хувирсан гаралтын хэлбэрт с а н г и й н утгыг буцаана. Хоосон тэмдэгтийг дүрмийн харилцан холбоос ба нөхцөлийн хэсэг хоёрт л ашигладаг.

Хоёр түвшинт дүрэм нь тэмдэгтийн цуваагаар илэрхийлэгдэх хэлийг бүрэн тодорхойлохын тулд нийтлэг тусгай тэмдэгт, операторыг ашигладаг. Тэдгээрийг хүснэгтээр харуулав.

***Хүснэгт 1.*** *Хоёр түвшинт дүрмийн тэмдэглэгээ*

|  |  |  |
| --- | --- | --- |
| № | Тэмдэглэгээ | Тайлбар |
| 1 | @ | ямар нэгэн тэмдэгт |
| 2 | : | харилцан холбоосын зааг |
| 3 | # | үгийн хязгаар |
| 4 | [] | тодорхой бүлгийг ялгах |
| 5 | {} | багана тус бүрээр ялгах |
| 6 | () | байж эсвэл байхгүй байж болно |
| 7 | | | эсвэл |
| 8 | + | үг залгаварын зааг |
| 9 | \* | хэдэн ч удаа давтаж болно |
| 10 | , | тэмдэгт зааглах |

* 1. **Төгсгөлөг төлөвт автомат**

Хоёр түвшинт дүрмийг хэрэглэхийн тулд хийсвэр хэлний төхөөрөмжийг ашиглан компьютерт хоёр түвшинт загварыг гүйцэтгэх боломжтой ба ийм үндсэн механик процедурыг төгсгөлөг төлөвт машин гэнэ. Хамгийн энгийн төгсгөлөг төлөвт машин бол төгсгөлөг төлөвт автомат ба тэр нь хэлний тэмдэгтүүдийг задлах, үүсгэх үйдлийг хийдэг. ТТА нь төлөв, шилжилт гэсэн үндсэн хоёр элементтэй. Төлөв нь эхлэлийн, төгсгөлийн, энгийн гэсэн гурван төрөл байдаг. Шилжилт нь нэг төлвөөс нөгөө төлөв үрүү шилжих шилжилтийг заана. Жишээ нь: цоож гэсэн тэмдэгтийн цувааг ТТА –аар илэрхийлбэл:

***Зураг 3.*** *цоож гэсэн тэмдэгтийн цувааны ТТА*

Төлөв 1 бол эхний төлөв ба төлөв 3 бол төгслийн төлөв болно. Төлөв нэг ба хоёр бол хоёулаа төгсгөлийн төлөв биш юм. Гурван шилжилттэй энгийн ТТА-ыг үзүүлсэн байна. Сумнууд нь шилжилтийг заана. ТТА нь мөн төгсгөлөг хувьсалын хүснэгтээр илэрхийлэгднэ. Дээрх бичвэрийг мөн хүснэгтээр илэрхийлбэл:

ц о ж

-------

1. 2 0 2

2. 3 2 3

3: 0 0 0

Хүснэгтийн мөрүүд ТТА-ын диаграммын гурван төлөвийг илэрхийлэх бөгөөд төгсгөлийн төлөвийн дугаар нь тодорхойлох цэгтэй ба төгсгөлийн бус төлвүүд нь цэгтэй байна. Баганууд нь нэг төлвөөс нөгөө төлөв үрүү шилжих сумыг илэрхийлнэ. ТТА-ын диаграм дах тэмдэгтийн түвшингийн сум бүр нэг баганын толгой болно. ТТА диаграмд ижил үсэг хоёр удаа орж, хоёр эсвэл нэг нумаар тэмдэглэгдсэн байхад хүснэгтэнд нэг баганаар илэрхийлэгднэ. Мөр ба баганын огтлолцолын цэгийн утгууд өөрчлөлт хийх төлвийг заадаг. Тухайн нүдэнд тэг байвал оролтын тэмдэгтийн төлөв ямар ч өөрчлөлт байхгүй буюу диаграмд тухайн төлөвт нум байхгүй гэсэн үг юм.

**ГУРАВДУГААР БҮЛЭГ МОНГОЛ ХЭЛНИЙ ҮГЗҮЙН ШИНЖИЛГЭЭ**

Монгол хэл нь түрэг, манж, халимаг хэлтэй нийлж Алтай язгуурын залгамал хэлний бүлгийг бүрдүүлдэг. Монгол хэл аялгуунд могол хэл, ойрд хэл, буриад хэл, монгол хэл, дуншаан хэл, боань хэл багтдаг. Монгол хэл аялгуугаар дэлхий дээр нийтдээ 8 сая хүн ярьдаг гэсэн тоо байдаг [3].

Хэл боловсруулалтын хөгжүүлэлтийг тухайн хэлний цахим бичгийн орчин, зөв бичгийн шалгуур, засуур, үгийн үүсгүүр, задлуур, өгүүлбэр зүйн задлуур, дүрмийн шалгуур, бичвэрийн боловсруулалт гэх мэтчилэн шат дараалан хийх шаардлагатай талаар өмнөх бүлэгт дурдсан.

Монгол хэлийг цахимд боловсруулах ажил нэлээд эрчимтэй явагдаж байгаа юм. МУИС-ийн “Компьютер хэл шинжлэлийн төв”-ийнхөн англи монгол машин орчуулгад өгөгдөлд суурилсан аргаар үгзүйг хувирган ашиглаж байгаа нь үр дүнтэй хэдий ч арай хялбар, хурдтай ажиллах боломжтой хоёр түвшинт загварыг монгол хэлний үгзүйд бүрэн туршиж үзэх шаардлагатай юм. Тиймээс эхлээд үгзүйн үүсгүүр, задлуурыг олон хэлэнд туршигдсан ПСи-Киммо хоёр түвшинт загвараар өмнө нь хийгдэж байсан судалгааны ажилд тулгуурлан, түүний дутагдалтай талыг засан, сайжруулж монгол хэлний үгзүйг бүрэн боловсруулалт хийхэд монгол хэлний үгзүйн хэв шинжийг нарийн тодорхойлж, хоёр түвшинт дүрэмд хэрэгжүүлэхэд бэлдэхээр ажилласан. Үүний тулд зөв бичгийн дүрмийг эхлээд судлан цааш үгзүйн боловсруулалт хийх болно.

**3.1 Орчин цагийн монгол хэлний зөв бичих зүй**

Монгол улсад 1940-өөд оны үеэс славян цагаан толгойд суурилсан шинэ бичиг үсэг хэрэглэх болжээ. Шинэ кирилл үсэг бол бичлэг дуудлага хоёрыг ихээхэн ойртуулж чадсан [2].

Монгол хэл кирилл тэмдэгт дээр суурилсан 35 тэмдэгт бүхий цагаан толгойтой (Зураг 1). Монгол цагаан толгой нь 13 эгшиг үсэг (а, э, и, о, у, ө, ү, я, е, ё, ю, й, ы), 20 гийгүүлэгч үсэг (м, н, г, л, б, в, р, ц, ж, з, с, д, т, ш, ч, х, п, к, ф, щ), 2 тэмдэг үсэг (ь,ъ) - ээс тогтдог.

Цагаан толгойн 35 үсэг

13 эгшиг үсэг

20 гийгүүлэгч үсэг

2 тэмдэг үсэг

Үндсэн 7 ýãøèã

Туслах 6 эгшиг

Эгшигт 7 гийгүүлэгч

Заримдаг 9 гийгүүлэгч

Онцгой гийгүүлэгч

а, э, и, о, у, ө, ү

я, е, ё, ю, й, ы

м, н, г, л, б, в, р

ц, ж, з, с, д, т, ш, ч, х

п, к, ф, щ

ü,ú

***Зураг 4.*** *Монгол хэлний цагаан толгой.*

Үндсэн 7-н эгшиг бол дангаараа хэрэглэгдэх ба гийгүүлэгч эгшигжүүлнэ. Туслах эгшгүүдээр гийгүүлэгч + эгшиг (j+э) гэсэн хоёр авиаг тэмдэглэдэг.

Эгшигт гийгүүлэгчид нь монгол үгэнд өмнөө буюу хойноо аль тод хэлэгдэх талдаа заавал эгшигтэй бичигддэг. Жишээлбэл: монгол, үйлдвэр, баатар, цуваа. Заримдаг гийгүүлэгчид нь үгийн дунд, адагт орохдоо заримдаа хойноо буюу урдаа эгшигтэй, заримдаа аль ч талдаа эгшиггүй бичигддэг. Жишээлбэл: танхим, хүрхрээ, алх. Онцгой гийгүүлэгчид нь гадаад үгсийг бичихэд хэрэглэгддэг. Жишээлбэл: фирм, пионер, карт. Тэмдэг үсэг нь ямар ч дуудлагагүй бөгөөд зарим гийгүүлэгчийг хэлэгдвэл зохих байрнаас нь урагшлуулах, хойшлуулах зэрэг үүрэгтэй. Тухайлбал я, ё хоёрыг өмнөх гийгүүлэгчээс нь саланги дуудахын тулд я, ё-гийн өмнө хатуугийн тэмдэг (ъ) бичнэ. Жишээлбэл: гаръя, уулзъя, очъё, олъё.

Монгол хэл нь дээрх цагаан толгойн үсгүүдэд дараах дүрмүүдийг баримтлан үгзүйн зөв бичих зүй болно.

**3.1.1 Эгшиг зохицох ёс**

Монгол хэлний нэг үгэнд янз бүрийн эгшиг хамт ордоггүй нарийн журам бий. Үгийн өргөлттэй үед а,о,у эгшиг байвал дараах үеүдэд э, ө, ү орохгүй, үгийн өргөлттэй үед о эгшиг байвал дараах үеүдэд а огт орохгүй мэт. Ингэж өргөлт бүхий нэгдүгээр үеийн эгшиг уул үгийн дараах үед байх эгшгийг тогтоох журмыг эгшиг зохицох ёс гэнэ. Энэ дүрмийг хүснэгтээр харуулав.

***Хүснэгт 2.*** *Эгшиг зохицох ёс.*

|  |  |
| --- | --- |
| Үгийн эхний үед байх эгшиг нь | Хойшид тохиолдох эгшиг нь |
| а  у | А |
| э  ү  и | Э |
| о | О |
| ө | ө |

Тайлбар:

1. Үеийг тэмдэглэсэн я, е, ё, ю үсгүүд өөрийн бүтцэд байгаа а, э, о, у, ө, ү эгшгээр зохицох буюу өөрөөр хэлбэл богино, урт, хос эгшгийг тэмдэглэсэн эдгээр үгсүүд богино, урт, хос а, э, о, у, ө, ү эгшгийн адил зохицно.

Жишээлбэл: явъя (я-гийн дараа а орж зохицсон)

ёотон (ё-гийн дараа о орж зохицсон)

Өргөлттэй үеэс хойш орсон уу, үү (юу, юү, яу, ёу, еу) зэрэгт урт эгшгүүд дараах үедээ орох эгшгийг өөрийн нөлөөнд оруулж дахин зохицуулдаг. Жишээлбэл: оруулах, өргүүлсэн, оюутан, гиюүрэл, хаяулдаг, гоёулав, биеүдэд.

## Өргөлттэй үеэс хойш орсон и, ий, эй, эгшгүүдэд дараах үедээ орох эгшгийг зохицуулахгүй, тэдгээр эгшиг нь өргөлттэй үеийн эгшгийг даган зохицсон хэвээр байна. Жишээлбэл: Тархинаас, өнгөтэйхөн.

## Гээгдэх эгшгийн дүрэм

Балархай эгшигтэй үгэнд зарим нөхцөл, дагавар залгахад балархай эгшиг нь хэрэглэгдэхгүй ёс бий. Ийм эгшгийг бичигт орхиж бичиж гээгдэх эгшиг гэж нэрлэдэг. Эгшиг гээгдэх тохиолдолууд:

1. Гийгүүлэгчээр төгссөн үгэнд урт эгшгээр эхэлсэн залгавар залгахад уул үгийн эцсийн гийгүүлэгчийн өмнөх балархай эгшиг зөв бичих дүрмийн бусад дүрэмд харьяалахгүй бол гээгдэнэ. Жишээлбэл: Утас + аар = утсаар

**2.** И-гээс бусад богино эгшгээр төгссөн үгэнд урт эгшгээр эхэлсэн залгавар залгахад эцсийн богино эгшиг гээгдэнэ. Жишээлбэл: Мөнгө + өөр = мөнгөөр

3. Гийгүүлэгчээр төгссөн үгэнд гийгүүлэгчээр эхэлсэн залгавар залгахад уг залгаврын өмнө эгшиг шаардагдвал шаардагдсан эгшгийг нь бичиж, эцсийн гийгүүлэгчийн өмнөх богино эгшиг гээгдэнэ. Жишээлбэл: Боловсор + л = боловсрол

4. Үгийн дунд орсон Ж, Ч, Ш-ийн дараах богино эгшиг И урт эгшгээр эхэлсэн залгавар залгахад гээгдэнэ. Жишээлбэл: түшиг + ээр = түшгээр

**3.1.3 Гээгдэхгүй эгшгийн дүрэм**

1. Эгшигт гийгүүлэгчийн аль нэг нь эгшиггүйдэхэд хүрвэл балархай эгшгийг гээж болохгүй. Жишээлбэл: Дүрслэл + ээр = дүрслэлээр

2. Заримдаг гийгүүлэгчийн дараа орсон заримдаг гийгүүлэгч эгшиггүйдэхэд хүрвэл балархай эгшгийг гээж болохгүй. Жишээлбэл: Нөхцөл + өөр = нөхцөлөөр

3. Хэлний үзүүрийн Н, хоолойн Г-ийн ялгах эгшгийг гээж болохгүй. Жишээлбэл: Санал + аар = саналаар, Хагар + аад = хагараад

4. Ж, Ч, Ш гурваас бусад гийгүүлэгчийн дараа орсон И эгшиг өмнөх гийгүүлэгчийн зөөлөрснийг тэмдэглэдэг учир гээж болохгүй. Жишээлбэл: Гуниг + аас = гунигаас, Авир + ын = авирын

5. Үйлт нэрийн ирээдүй цагийн Х нөхцөлийн өмнөх балархай эгшгийг гээж болохгүй. Жишээлбэл: давтах + ыг = давтахыг, үзэх + ээс = үзэхээс

**3.1.4 Б, В гийгүүлэгчийг зөв бичих**

1. Үгийн дунд ба адагт В, Л, М, Н дөрвөн гийгүүлэгчийн дараа зөвхөн Б гийгүүлэгчийг бичнэ. Жишээлбэл: Самбар, салбар, авбал

2. В, Л, М, Н гийгүүлэгчээс бусад бүх гийгүүлэгч болон эгшгийн ард В бичнэ. Жишээлбэл: жолоодвол, маллавал

## 3.1.5 Зөөлний тэмдгийг И эгшгээр солих дүрэм

## 1. Зөөлний тэмдгийн дараа а, о, у, и эгшгийн аль нэгээр эхэлсэн дагавар, нөхцөл орвол уг зөөлний тэмдгийг и-ээр сольж бичнэ.

## 2. Зөөлний тэмдгийн дараа эгшигт гийгүүлэгчийн аль нэг буюу түүгээр эхэлсэн дагавар, нөхцөл орвол зөөлний тэмдгийг И болгон бичнэ.

## 3. Үйлт нэрийн х дагавар ямагт өмнөө эгшигтэй бичигддэг тул энэхүү х-гийн өмнөх зөөлний тэмдгийг и болгоно. Жишээлбэл:

## сахь+ж-сахиж, захь+дал-захидал

## 3.1.6 Зөөлний тэмдгийг И эгшгээр солихгүй тохиолдол

## 1. Зөөлөрсөн эгшигт гийгүүлэгчийн дараа заримдаг гийгүүлэгч орвол зөөлний тэмдгийг хэвээр бичнэ. Жишээлбэл: хонь+чин-хоньчин, хоньд, морьтой тарь+сан-тарьсан гэх мэт

## 2. Зөөлний тэмдгийн дараа туслах эгшиг орвол зөөлний тэмдэг И болохгүй. Жишээлбэл:барья,сорьё,сахья,больё г.м

## 3.1.7 Эгшигт гийгүүлэгчийн дүрэм

## М, Н, Г, Л, Б, В, Р-энэ 7 гийгүүлэгч өмнөө буюу хойноо аль тод хэлэгдэх талдаа заавал эгшигтэй бичигдэнэ. Жишээлбэл: сурагч, хангинана

## Зөрчих тохиолдол: Хэлний угийн Н-ийн ард заахын тийн ялгалын Г эгшиггүй бичигдэнэ. Жишээлбэл: дүнг, анг, хадланг

## 3.1.8 Заримдаг гийгүүлэгчийн дүрэм

## 1. Д, Т, С, Х, Ц, Ч, Ш, З, Ж-энэ 9 заримдаг гийгүүлэгч нь үгийн дунд, адагт орохдоо заримдаа хойноо буюу ардаа эгшигтэй, заримдаа аль ч талдаа эгшиггүй бичигддэг. Жишээлбэл: төрөлхтний, танхим, хүндлэл

## 2. Заримдаг гийгүүлэгч нь Н, Г, Л, М, В, Р, Б гийгүүлэгчийн дараа орсон тохиолдолд эгшиггүй бичигдэнэ. Жишээлбэл: Ханш, ялалт, идэвч

## 3.1.9 Дараалан орсон гурван гийгүүлэгчийн дүрэм

## 1. Дараалан орсон хоёр гийгүүлэгчийн хоёрдугаар нь хоолойн Г, хэлний үзүүрийн Н байвал өмнө нь эгшиг жийрэглэхгүй. Жишээлбэл: алга,бичнэ,олно

## 2. Дараалан орсон гурван гийгүүлэгчийн гуравдугаар гийгүүлэгч нь хоолойн Г, хэлний үзүүрийн Н байвал Н, Г-ийн өмнө эгшиг жийрэглэнэ. Жишээлбэл: бодогд-(о)-но, янзага, давш-(и)-на

## 3.1.10 Н, Г –гийн дүрэм

## 1. Бичигт Н үсгээр хэлний үзүүрийн Н, угийн Н гэсэн 2 авиаг тэмдэглэдэг. Хэлний угийн Н ардаа эгшиггүй, үзүүрийн Н ардаа эгшигтэй бичигддэг. Жишээлбэл: сана-сан, хана-хан, оно-он

## 2. Г үсгээр хоолойн Г, хэлний Г хоёр авиаг тэмдэглэдэг. Хоолойн Г-гийн дараа а, о, у, ы дөрвөн эгшгийн аль нэгийг бичнэ. Харин хэлний Г-г огт эгшиггүй бичих буюу а, о, у, ы дөрвөөс бусад эгшиг бичиж тэмдэглэнэ. Жишээлбэл: хорго хөөрөг, дуган тэнэгэр, шударга ингэ

## 3.1.11 Ялгах эгшгийн дүрэм

## Үгийн нэгдүгээр үеэс хойш хэлний үзүүрийн Н-ийн яг ард нь орсон а, э, о, ө эгшгийг, мөн нэгдүгээр үеэс хойш хоолойн Г-ийн ард бичигдсэн а, о эгшгийг ялгах эгшиг гэнэ. Жишээлбэл: Тэнэгэр, эрдэнэ, ургана, тонгочно

## 3.1.12 Ж, Ч, Ш-ийн дүрэм

## 1. Үгийн эхэнд орсон Ж, Ч, Ш-ийн дараа эр үгэнд а, о, у, эм үгэнд и, ү, ө гурвын аль сонстохыг бичнэ. Харин үгийн эхэнд орсон Ж, Ч, Ш-ийн дараа эр үгэнд и, эм үгэнд э эгшиг огт бичигддэггүй. Жишээлбэл: Жимс, шир, чавх

## 2. Үгийн дунд орсон Ж, Ч, Ш-ийн дараа богино эгшиг шаардагдвал эм, эр аль ч үгэнд зөвхөн и бичнэ. Жишээлбэл: ажил, догшин

## 3.1.13 Эр, Эм үгийн ИЙ, Ы нөхцөлийг зөв бичих дүрэм

1. Эр эгшгээс бүтсэн үгийг эр үг гэнэ. Эр үгэнд Ы нөхцөл бичнэ. Жишээлбэл: малгай-ны, араан-ы, ноолуур-ын

2. Эм эгшгээс бүтсэн үгийг эм үг гэдэг. Эм үгэнд ИЙ нөхцөлийг бичнэ. Саармаг эгшиг И-гээс бүтсэн үгэнд эм үгийн ИЙ нөхцөлийг бичнэ. Жишээлбэл: эмээл-ийн, элс-ний, илч-ийг

3. Хэлний Г буюу Ж, Ч, Ш-ээр төгссөн эр үгэнд эм үгийн ИЙ-г бичдэг. Жишээлбэл: хайч-ийг, буурч-ийн, ара(а->0)г-ийг, багаж-ийг

Эдгээр дүрмийг шууд хоёр түвшинт дүрэмд хөрвүүлэх бичих боломжгүй ба нэг дүрэм нь нөгөөгөө зарим талаараа үгүйсгэсэн, гажсан тохиолдлууд олонтойн дээр үг үүсгэх бүтээвэр буюу дагавар, нөхцөлийн хэлбэрээс ихээхэн хамаарах нь харагдаж байна. Иймд монгол хэлний бүтээвэр зүйг судлах хэрэгтэй.

**3.2 Монгол хэлний бүтээвэр зүй**

Монгол хэлний үгзүйн бүтцийг зөв ойлгохын тулд аливаа хэлний үгийн бүтцийн үндсэн нэгж болох бүтээврийн тухай товч авч үзье. Бүтээвэр нь хэлний шаталсан тогтолцооны биеэ даасан түвшний нэгж юм. Бүтээвэр түүний байрлал, бүтэц, утга, үүргийг судлах салбар ухааныг бүтээвэрзүй (морфемика) гэдэг. Өргөн утгаар бүтээвэр гэдэг нь хэлний тогтолцооны хамгийн бага утгат нэгж бөгөөд түүнд үгийн язгуур, дагавар, нөхцөл, угтвар үг, сул үг, ялгац зэрэг хэлзүйн үүрэгтэй туслах хэсгийг багтаадаг. Бүтээвэр нь үе авиалбараар биежиж, үг бүтээх ба хувилгах үүрэг гүйцэтгэдэг. Монгол хэл нь угтвар, оруулбар байхгүй язгуураас хойш нэг чигийн харьцаанд үг нь бүтэж, хувилдаг залгамал хэл билээ. Монгол үгийн бүтээврүүд нь язгуур, дагавар, нөхцөл гэсэн ерөнхий дараалалтай. Энэхүү дарааллыг зургаар харуулав.

***Зураг 5.*** *Монгол хэлний үг үүсгэх арга*

Монгол хэлний бүтээврүүд нь ихэвчлэн ганц утгатай. Жишээлбэл: унш-(и)-в гэсэн үгийн в нөхцөлөөр зөвхөн өнгөрсөн цагийн утга илэрч байна. Монгол үгийн бүтцэд хэлзүйн ба үгийн сангийн утга нь ил бүтээврээр илрэхээс гадна хоосноор илэрч болно. Ийм бүтээврийг тэг (∅) бүтэээвэр гэнэ.

Бидний судалгааны зорилгод аль болох хувилах үг буюу язгуур бүтээврээс хамаарахгүйгээр дагавар, нөхцөлөөр зөв хувиргахаар зорьсон. Иймд язгуур бүтээврийг орхин, харин хоёр түвшинт дүрэмд тохируулан дээрх зорилгын дагуу бүх дагавар ба нөхцөлийг түүн, нэг төрлийн нөхцөлийг ерөнхийлөн тэмдэглэх, зөв кодлох шаардлагатай гэж үзсэн.

Үгийн язгуур ба үндсэнд залгаж шинэ үг бүтээх залгавар бүтээврийг дагавар, үгийн язгуур ба үндсэнд залгаж шинэ үг бүтээхгүй, зөвхөн уул үгээ хувилгах, хийсвэр харьцаа заах үүрэгтэй залгавар бүтээврийг нөхцөл гэнэ. Үг хувилгах залгаварт нэрийн тийн ялгал, хамаатуулах, олон тоо, үйл үгийн цаг, төлөв байдал, тодотгон холбох болон нөхцөлдүүлэн холбох нөхцөлийг оруулдаг. Нөхцөл нь хийсвэр утгатай боловч нэг нөхцөлийг маш олон үндсэнд (холбогдох үгсийн аймгийн бүх үндсэнд) залгаж болдог. Тэгвэл нэг дагаврыг цөөн хэдэн язгуур, үндсэнд залгана.

**3.2.1 Монгол хэлний дагавар, нөхцөлийг хоёр түвшинд дүрэмд бэлдэн ангилах**

Эл чиглэлийн өмнөх судалгаануудаас илүү сайжруулахын тулд хувирах үг буюу язгуураас үл хамааран дагавар, нөхцөлийг залгахад зөв хувиргадаг байх шаардлагатай гэж дээр дурдсан. Иймээс дагавар, нөхцөлийг нэгдүгээрт үгийн үүсгүүрт зориулж ерөнхийлөн нэг үсгээр тэмдэглэх, хоёрдугаарт үгийн задлуурт зориулж нөхцөл бүрийг кодлох хэрэгтэй гэж үзлээ.

Энэ нь нэг төрлийн нөхцөл тухайн үгнээсээ хамаарч хэд хэдэн хэлбэрийг үүсгэдэг. Жишээ нь:

*Тооны нөхцөл:*

* *ууд, үүд салбарууд, дэвтрүүд*

*Тийн ялгал:*

* *Харьяалах: ийн, ын, ий, ы, н, ны ахын, эгчийн, хэлний, орны*

*Хэвийн нөхцөлүүд:*

* *лга, лгэ, лго, лгө угаа-лга, хий-лгэ, гөлий-лгө, ороо*

Дээрх жишээнээс харахад харъяалахын тийн ялгалын “ийн” гэж дуудагдах нэг төрлийн нөхцөл тухайн хувирах үгнээсээ хамаарч зургаан хэлбэртэй байна. Үүнийг ерөнхийд нь нэг тэмдэглэгээгээр “Ын” тэмдэглэн оруулж, хоёр түвшинт дүрмээ тухайн үг ямар нөхцөл авахыг таниж чадахаар зохион дээрх зургаан хэлбэрт шилжинэ.

Залгавар бүтээврүүдийг хэрхэн тэмдэглэснийг доорх хүснэгтэд заримыг харуулав. Бүх залгавар бүтээврийн тэмдэглэгээг хавсралт А-д хавсаргав.

***Хүснэгт 3.*** *Залгавар бүтээврийн тэмдэглэгээний жишээ*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Зохицох эгшиг  Бүтээвэр | ө | э | а | о | и | ы | у | ү |
| ийн, ын, ий, ы, ны (Ын) |  |  |  |  | + | + |  |  |
| ийг, ыг, г (Ыг) |  |  |  |  | + | + |  |  |
| маг, мог (мОг) |  |  | + | + |  |  |  |  |
| уул, үүл (УУл) |  |  |  |  |  |  | + | + |
| аас, ээс, оос, өөс (ҮҮс) | + | + | + | + |  |  |  |  |
| мад, мэд (мАд) |  | + | + |  |  |  |  |  |
| мхай, мхий (мхИй) |  |  | + |  | + |  |  |  |

Дагавар, нөхцөлийн төрлийг кодлосноор үгийн задлуурт ямар нөхцөл, дагавраар хувирсныг таних ба үүнийг олон хэлний кодын схемийн кодлох зарчмын дагуу кодолсон. Кодын схем нь хоёр түвшинт дүрмийн оруулах, устгах үйлдлийг танин дагавар нөхцөлийн кодоор ялгадаг.

Кодын загвар P0 P1 P2 P3……\_\_\_ xyz

P0 : Үгийн аймгийн үгзүйн категори код

Xyz: Дагавар нөхцөлийн код

Монгол хэлний үгийн аймгийн үгзүйн категори кодууд:

Дагаврыг шууд байгаа хэлбэрээр нь кодлосон ба нөхцөлийн хувьд:

***Нэрийн хувилал***

Нэр үг нь гурван төрлийн нөхцөлөөр хувирдаг. Үүнийг:

***Үйл үгийн хувилал***

Үйл үг дараах нөхцөлүүдээр хувилна.

Ингээд монгол хэлний бүх дагавар нөхцөлийг дараах байдлаар ерөнхийлөн тэмдэглэж, кодлон, ангиллаа. (Хавсралт А).

Монгол хэлний үгзүйн шинжилгээний явцад дагавар, нөхцөлийн тоо олон, үгийн язгуурууд өөр хоорондоо маш уян хатан, дээрх зөв бичих дүрэмд бүрэн захирагдахгүй, эсвэл онцгой байдлаар залгагддаг нь харагдаж байсан бөгөөд энэ нь монгол хэлний үгзүйг цахимжуулахад хүндрэлтэй, цахимжуулалтын туршлагыг авч болох бусад төрлийн хэлнээс ялгагдаж байгаа гол судалгаа шинжилгээний хэсэг юм.

Монгол хэлэнд үеийг ар араас нь залган үгийг бүтээх бөгөөд тэдгээр үеийг холбох ямар ч түвшинд гийгүүлэгчээс шалтгаалж эгшиг (зарим тохиолдолд эгшгээс шалтгаалж гийгүүлэгч ) жийрэглэх, гээх, солих үзэгдлүүд ижил илэрдэг. Шинэ үсгийн дүрмийг олон нийтээр зөвшөөрөн хэрэглэсээр 60 гаруй жил болж байна. Гэвч энэ дүрмийг шүүмжлэх явдал мөн хугацаанд гарсаар байгаа. Огцом богино эгшгийг бичнэ, гээгдэх эгшгийг гээхгүй болгоно, үгзүйн зарчмыг яг баримтална гэх зэргийн олон шүүмжлэл гарсаар байдаг нь монгол хэлийг яг бүрэн дүрмээр зохицуулагдахад бэрхшээлтэй нь харагдаж байна.

Тиймээс монгол хэлний үгзүйн дүрэм, бүтээврийг судалсны үндсэн дээр хоёр түвшинт дүрмээ зохиох буюу цахим үүсгүүр, задлуурыг үүсгэх нэмэлт дүрмийг язгуурын төгсгөл, дагаврын эхлэл хоёрыг тогтоож, нэгтгэн залгахаар зохиох хэрэгтэй юм.

**ДӨРӨВДҮГЭЭР БҮЛЭГ ХОЁР ТҮВШИНТ ДҮРЭМ БОЛОВСРУУЛАЛТ, ҮР ДҮН**

Хүнд зориулсан хэлний судалгаа нь ерөнхий тогтолцоог илэрхийлсэн мэдлэг байдаг. Учир нь, хүн тэрхүү мэдлэгтээ түшиглэн, бусад тохиолдлыг тогтолцоогоор нь үүсгэн бүтээж, чөлөөтэй сэтгэж чаддаг. Харин хиймэл оюун ухаан бүхий цахимд зориулсан “хүмүүний хэл” –ний судалгаанд үйлдэл бүрийг нарийн тодорхойлох хэрэгтэй. Тэгвэл цахимд хөрвүүлэх програмын хувьд хэрэглэгчийн өгч буй орох бичвэр (input text)-ийг задлаад, боловсруулж, дагавар, нөхцөлийг залган зөв хувирган гарах бичвэр (output text) –ийг үзүүлэх ёстой тул *хоёр тийш чиглэсэн* мэдлэг байх ёстой юм. Тэгэхээр хүний ухамсарт *“үгийн язгуур үндэс ”* эхлэл болдог бол, цахимын хувьд *“хамгийн сүүлийн үсэг буюу тэмдэгт”* нь боловсруулалтын эхлэх цэг болдог.

Үгийн төгсгөлөөс урагш чиглэсэн, хүний ухамсрын гүйцэтгэдгээс эсрэг чиглэсэн үйлдэл болох эхний алхамд буюу *“үгзүйн задлуур”* (morphological analyzer) –т дараах үйлдлийг хийнэ.

* Орох бичвэр дэх үгсийн төгсгөлийн тэмдэгтээс эхлэн шинжилж байж болох залгаварын хэлбэрийг онож тогтоох
* Тухайн залгаврын өмнө ямар залгавар орсон болохыг тогтоох
* Төгсгөлийн залгавар болон түүний өмнөх залгаврын хослол байж болох эсэхийг шалгах
* Шат шатаар задлан үгийн үндэс хүртэл задалж, үгийн үндсийг таних
* Залгаврын дарааллыг хадгалах

Үгийн үндсээс төгсгөл рүү чиглэсэн хоёр дахь алхам буюу “үгзүйн үүсгүүр”(morphological generator) –т дараах үйлдлийг хийнэ.

* Орох бичвэрт үгийн сангийн хэлбэрээр өгөгдсөн үгийг дагавар, нөхцөлөөр зөв хувилж үгийн хувирсан хэлбэрийг үүсгэдэг. Ингэхдээ оролтын үгийг төгсгөлөг төлөвт автоматаар тохирох дүрмийг олж хувиргадаг.

Хоёр түвшинт дүрмийг хэрэгжүүлэх ПСи-Киммо-ийн ажиллагааны бүдүүвч зургаар илэрхийлбэл дараах байдалтай байна [4]. (Зураг 6)

Two level rules

LEVEL

RULES

**RESOURCES**

**PC-KIMMO**

**RULES**

**LEXICONS**

RECONIZER

RECONIZER

GENERATOR

GENERATOR

KimmoData

KimmoData

KimmoResult

KimmoResult

KimmoResult

KimmoResult

**харагдана**

Surface form

**харагд + нҮ**

Lexical form

**V(харагд) +VVI\_ ETFUT**

Gloss

**харагдана**

Surface form

**KGEN**

**KGEN**

**PC-KIMMO** Regular Expressions

LOADS

**Зураг 6**. ПСи-Киммо-ийн ажиллагааны бүдүүвч

Зураг 6-д үзүүлсэн KGEN нь хоёр түвшинт илэрхийллээр бичсэн зөв бичих дүрмийг хоёр түвшинт үг зүйн дүрэмд хөрвүүлдэг дэд програм юм. KimmoData нь ПСи-Киммо програмд хэрэгтэй нөөцүүдийг агуулдаг бол KimmoResult нь үүсгүүр болон задлуурын үр дүнг агуулна. ПСи-Киммо 2 хувилбарт Шайберийн PATR II загварчлалд суурилсан үгийн задлалын мод зурагч шинээр орж ирсэн [5]. Энэ нь олон нөхцөл, дагавар дараалан орж хувирсан үгийг задлах задлуурт хэрэглэгдэх үгзүйн дарааллын файл (mon.grm) юм. Үүнийг цаашид судлан, нэмэх боломжтой юм.

Хоёр түвшинт илэрхийллээр бичсэн зөв бичих дүрэм болон үгийн сангийн файлыг хэрэглэгч гараараа бэлдэж өгөх шаардлагатай байдаг.

**4.1 Хоёр түвшинт дүрэм зохиох**

Хоёр түвшинт дүрмийг зохиох процессын үе шатуудад хоёр зарчмыг баримтлан гүйцэтгэх ёстой. Нэгд, ерөнхий зорилгодоо хүрэхийн тулд тухайн хэлний бүх дүрмийг оруулах, Хоёрт, дүрэмд хамааралтай тухайн хэлний дүрэмд зөрчилдөхгүй байх боломжит хувилбарыг оруулахыг хичээх.

Хоёр түвшинт дүрэм үүсгэх гурван хэлбэрийг танилцуулъя. Эхний арга нь өгсөн хэлний үгзүйн дүрмийг ашиглан түүнийгээ хоёр түвшинт дүрэмд шилжүүлэх. Хэрэв тэдгээр дүрэмд биелэхгүй өгөгдөл олдвол түүний учрыг тайлбарлах хэрэгтэй. Үүний шалтгаан нь гадаад үг хэллэг, дүрмийн гажилттай үгс орно. Хоёр дахь хэлбэр нь үгийн сангийн болон хувирсан хэлбэрийг ашиглан хоёр тогтсон тэмдэглэгээг хувирган зөв шийдэлд хүрэх. Гурав дахь нь янз бүрийн нөхцөл байдалд өгөгдсөн үгсийг сайтар судлаад тэдгээр хэлбэрүүдийг харьцуулах замаар далд буй дүрмийг олж авахыг оролдоно [5].

Бид хувирах утгат хэсгийг дараах байдлаар тэмдэглэгээндээ ашигласан.

1. Б: б, в

2. Д: д, т

3. Ж: ж, ч

4. Ы: ий, ы

5. А: а, э

6. Э: а, о

7. И: а, и

8. У: у, ү

9. Й: а, э, о

10. Ө: а, и, о

11. Ү: а, э, о, ө

Дүрэм зохиоход үгийн сангийн өгсөн нэг үсэг хоёр буюу түүнээс олон хувирсан хэлбэртэй байж болно. Мөн үүнтэй холбогдолтой тодорхой зарчмуудыг гуравдугаар бүлэгт дэлгэрэнгүй тайлбарласан билээ.

* + 1. **Эгшиг зохицол**

Монгол хэлэнд эгшиг зохицох ёс бол зөв бичих дүрмийн гол түлхүүр болдог ба бүх үгэнд хэрэгждэг дүрмийн томоохон хэсэг нь юм. Үүний бичгийн дүрмийг 3.1.1 хэсэгт дэлгэрэнгүй өгүүлсэн. Өмнөх судалгаануудад энэхүү эгшиг зохицох ёс тусгагдаагүй юм. Тэгвэл үүнийг хоёр түвшинт дүрэмд хэрэгжүүлэхийн тулд Ү: а, э, о, ө ерөнхий тэмдэглэгээгээр тэмдэглэсэн. Энэ нь дээрх дөрвөн эгшиг орсон дагавар, нөхцөлийг оролтонд ерөнхий тэмдэглэгээр бичнэ. Харин бидний зохиосон эгшиг зохицох ёсны дүрэм нь үгийн язгуур үндэсийн төгсгөлийн эгшгийг ямар байгаагаас хамааран эгшиг зохицох ёс дүрмийн дагуу аль тохирох эгшгийг сонгон хувирна. (Хүснэгт 2) Хүснэгтээс харахад а ба э болох тохиолдлууд нэгээс олон, ялгамжтай байгаа тул тус бүр дэд олонлог болгон зарлаж, энгийн дүрэм зохиолоо. Жишээ нь: С- гийгүүлэгчийн дэд олонлог, \* -тэг эсвэл нэгээс олон байж болно, | - нь а эсвэл у байж болно, [] эсвэл үйлдлийг ялгах хаалт, + - үг ба залгаврын зааг ба харилцан холбоост оролцож болно, @ - ямар нэгэн байдлаар хувирсан ба хувираагүй тэмдэгт байж болно, \_\_\_ - дүрэм зохиогдож буй харилцан холбоосыг илэрхийлнэ.

RULE Ү:а => [а|у] C\* +:@ C\* \_\_\_

RULE Ү:э => [э|ү|и] C\* +:@ C\* \_\_\_

RULE Ү:{о,ө} => {о,ө} C\* +:0 C\* \_\_\_

Хоёр түвшингээр илэрхийлбэл: **ҮТ:**Үгийн сангийн түвшин, **ХТ:**Хувирсан түвшин

**ҮТ:** г а р + н Ү гар+VVI\_ETFUT

**ХТ:** г а р 0 н а гарна

**ҮТ:** ү з + м Ү г ц үз+VVI\_CSPREP

**ХТ:** ү з 0 м э г ц үзмэгц

**ҮТ:** о р + д Ү г ор+VVI\_CDPRD

**ХТ:** о р 0 д о г ордог

**ҮТ:** ө г + л Ү г өг+NVD\_LAG

**ХТ:** ө г 0 л ө г өглөг

Гэтэл дээрх дүрмийг зохиоход энгийн, хялбар байгаа хэдий ч тухайн дүрэм нь “й” –ээр төгссөн (гүй+нҮ), сүүлийн эгшиг нь и боловч өмнөх эгшиг нь и-ээс бусад эгшиг орсон (нарги, дорги), туслах эгшиг орсон (яв, ёс), зөөлний тэмдэгээр төгссөн, балархай эгшиг нь гээгддэг, гээгддэгүй байх гэх мэтчилэн үгэнд хувиргахгүй ба зөв хувиргаж чадахгүй юм. Тиймээс энэхүү дүрмийг дээрх бүгд, мөн монгол хэлний бүх төрлийн үгэнд зөв зохицон хувирдаг нарийн тодорхой дүрэм зохиох түүнээс үүдэлтэйгээр нэмэлт олон дүрмүүд зохиогдох шаардлагатай болсон. Тэгэхээр эхлээд монгол хэлэнд нийт хэдэн төрлийн, ямар язгуур үндэс байдаг хэрхэн хувирдаг талаар авч үзэх хэрэгтэй болж байна. Үүнийг судалсны үндсэн дээр Ц.Дамдинсүрэн, Б.Осор нарын “монгол үсгийн дүрмийн толь”-д бичсэн нэрийн 29 төрөл, үйлийн 18 төрлийн үгсийг туршилтандаа ашиглан тэр бүгдэд хэрэгжих эгшиг зохицох дүрмийг зохиох юм.

Эндээс эгшиг зохицлыг судлан боломжит хослолуудыг гарган хүснэгтээр илэрхийлэв.

***Хүснэгт 4.*** *Эгшиг зохицох харилцан холбоос*

|  |  |  |
| --- | --- | --- |
| Эхний үе | Дэд олонлог | Дараагийн үед орох эгшиг |
| {а, аа, ай я, яа} | OO | Ү : а |
| {у, уу, уй, юу} |
| {э, ээ, эй} | FF | Ү : э |
| {ү, үү, үй, юү} |
| {и, ий} |
| {ө, өө, еө} |  | Ү : ө |
| {о, оо, ой, ё, ёо} |  | Ү : о |

Хүснэгтэд харилцан холбоос бүрд дээрх төрлийн үгнүүдэд тохирох дүрмийг зохиосон хэдий дахин хоёр асуудал тулгарсан. Үүнд урт эгшиг эсвэл хоёр том үсгээр тэмдэглэсэн тэмдэглэгээ давхардаж, алгасаж орох тохиолдол монгол хэлэнд маш өргөн хэрэглэгддэг. Жишээлбэл:

**ҮТ:** я в + Ү Ү р Ү й яв+VVI\_EP2ORDR

я в + а Ү р Ү й

я в + а а р Ү й

я в + а а р а й

**ХТ:** я в 0 а а р а й яваарай

**ҮТ:** х ө т ө л + Б Ү р хөтөл+NVD\_BAR

х ө т ө л + Б ө р

х ө т ө л + б ө р

**ХТ:** х ө т ө л 0 б ө р хөтөлбөр

Дээрх жишээнд гурван эгшиг орсон, мөн Б гийгүүлэгчийн тэмдэглэгээний б эсвэл в болж хувирах зэрэг орсон байна. Тэгэхээр эгшиг зохицох ёсны хоёр түвшинт дүрэм нь дээрх бүгдийг нэгтгэсэн нэлээд өргөн хүрээг хамарсан олон нэгдэл, харилцаа хамааралтай, их ажиллагаа шаардсан дүрэм зохиогдох нь харагдаж байна. Энэхүү эгшиг зохицох ёсны дүрмийг нэгтгэн харуулбал доорх хэлбэртэй байна.

RULE Ү:а => OO [и\* C\*|й C\*|ь: (ь:@)|а:(а:@) C\*] +:@ C\* и\* (Б: (Б:@)) (Ү: (Ү:@)) C\* \_\_\_

RULE Ү:э => FF [и\* C\*|й C\*|ь: (ь:@)|э: (э:@) C\*] +:@ C\* и\* (Б: (Б:@)) (Ү: (Ү:@)) C\* \_\_\_

RULE Ү:о => о [и\* C\*|й C\*|ь: (ь:@)|о: (о:@) C\*] +:@ C\* и\* (Б: (Б:@)) (Ү: (Ү:@)) C\* \_\_\_

RULE Ү:ө => ө [и\* C\*|й C\*|ь: (ь:@)|ө: (ө:@) C\*] +:@ C\* и\* (Б: (Б:@)) (Ү: (Ү:@)) C\* \_\_\_

Эгшиг зохицох ёсны эдгээр дүрмээс үүдэлтэй гээгдэх ба гээгдэхгүй байх тохиолдол, зөөлний тэмдгийн хувирал зэргийг дараагийн хэсгүүдэд тайлбарлая. Тэгэхээр хоорондоо холбоо хамааралтай дүрмүүд нэгэн зэрэг хэрэгжсэнээр эцсийн зөв хувирсан үгийг гаргана.

**4.1.2 Тийн ялгалын нөхцөлийн эгшиг зохицол**

Тийн ялгалын нөхцөлийг тусгайлан авч үзсэн нь монгол хэлний нэрийн бүх үг мөн үйлийн хувирсан хэлбэрүүд хувирдаг өргөн хүрээний нөхцөлүүд багтдаг. Монгол хэлэнд одоо 7-н тийн ялгалын нөхцөл байна. Мөн энд олон тооны нөхцөлийг хамт дурдъя. Тэдгээрийн хоёр түвшинт дүрмийн ерөнхий тэмдэглэгээг үзүүлэв.

***Хүснэгт 5.*** *Тийн ялгалын нөхцөлийн ерөнхий тэмдэглэгээ*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Нэршил | Нөхцөл, түүний төрөл | ХТД тэмдэглэгээ |
| 1 | Нэрлэх | 0 | 0 |
| 2 | Харьяалах | ын, ийн, н, ны, ы, ний | Ын, нЫ |
| 3 | Өгөх, орших | д, т | Д, нД |
| 4 | Заах | ыг, ийг, г, | Ыг |
| 5 | Гарах | аас, ээс, оос, өөс | ҮҮс, нҮҮс |
| 6 | Үйлдэх | аар, ээр, оор, өөр | ҮҮр |
| 7 | Хамтрах | тай, тэй, той | тЙй |
| 8 | Чиглэх | руу, рүү | рУУ |
| 9 | Олон тоо | ууд, үүд | УУд |
| чууд, чүүд | чУУд |

Дээрх тэмдэглэгээнүүдийн хувирах нөхцөлийн төрлийг зааж өгөн, монгол хэлний зөв бичгийн дүрмийн дагуу нэрийн бүх 29-н төрөлд зөв хувиргах, зөрчил үүсэхгүй байхаар мөн хоёр түвшинт дүрмийг зохиох хэрэгтэй.

Энд гарах, үйлдэхийн тийн ялгалын нөхцөлүүд өмнөх хэсгийн эгшиг зохицох ёсны дүрмийн дагуу хувирна.

Өгөх оршихын нөхцөлийн дүрэм нь {г, с, р, в} (SUBSET DP ) дараа “т” болно бусад бүх тохиолдолд “д” болох ерөнхий зарчмыг дагадаг ч үүнийг зөрчиж гаждаг тохиолдлууд бас байсан юм. (сард, хосод) Тэдгээр үгсийг толь бичгийн аргаар дүрэмжүүлэхээр шийдвэрлэсэн.

RULE Д:т => DP +:0 \_\_\_

Хамтрах, чиглэх, олон тооны нөхцөлүүд ерөнхий зарчмаараа хувирах нөхцөлийн төрлүүдээр задлан зохиогдсон.

Харьяалах, заахын тийн ялгал нь 3.1.13 дугаар зөв бичих дүрмийн дагуу бичигдэх хэдий ч түүнээс өөр олон дүрмүүд нэмэгдэж байж гэмээн нэрийн тэдгээр бүх төрлүүдэд хүчин төгөлдөр болох юм. Үүнд ж, ч, ш, г (SUBSET TT )–ийн дараа эр эм үг аль ч тохиолдол харгалзахгүй ийн, ийг нөхцөлөөр хувирна. Үүнийг хэрэгжүүлэхийн тулд хориглох дүрэм, өөр бусад олон хүчин зүйлийг тооцон зохионо.

RULE Ы:й <=> [[GG\* C\* |RR\* C\* TT|г |и C\*|C\* и: (и:@) QL] +:и|C\* и +:0|ь: (+:0)|[GG|GG C\*] +:0 0: (0:@) н: (н:@)] \_\_\_

RULE Ы:ы /<= [GG | GG C\*| RR\* C\* TT| г|аа|оо|уу] +:0 \_\_\_

**ҮТ:** б а г ш + Ы н багш+NNI\_GENI

**ХТ:** б а г ш и й н багшийн

Мөн энэхүү хоёр тийн ялгалын нөхцөл тэр дундаа харьяалахын нөхцөл нь нэрийн 29 төрөл бүр дээр өөр өөр зарчмаар хувирч эгшиг, гийгүүлэгч гээх, жийрэглэх, тэдгээрийг зөрчих тохиолдол гэх мэтчилэн маш олон өөрчлөлттэй тул түүнийг нэгтгэхэд нэлээд хүндрэлтэй нэмэлт дүрмүүд зохиосон. Жишээлбэл:

**ҮТ:** о ю у т а н + Ы н оюутан+NNI\_GENI

**ХТ:** о ю у т 0 н 0 ы 0 оюутны

**ҮТ:** т ө р ө л х т ө н + Ы н төрөлхтөн+NNI\_GENI

**ХТ:** т ө р ө л х т 0 н и й 0 төрөлхтний

**ҮТ:** б о р о о + Ы н бороо+NNI\_GENI

**ХТ:** б о р о о н 0 ы борооны

Харин харъяалахын тийн ялгалыг хоёр ерөнхий тэмдэглэгээгээр тэмдэглсэн нь тухайн нөхцөлийн үед монгол хэлний тогтворгүй Н авиа орж ирдэг. Тэгэхээр Н авиа гарч ирдэг үгийг тодорхойлох боломжгүй юм.

**ҮТ:** д а р г а + Ы н дарга+NNI\_GENI

**ХТ:** д а р г 0 0 ы н даргын

**ҮТ:** ч а р г а + н Ы чарга+NNI\_GENI

**ХТ:** ч а р г а 0 н ы чарганы

Монгол хэлний тогтворгүй Н авиа бол толь ба дүрмийг цэгцлэг боловсруулахад нэлээд бэрхшээлийг учруулж байна. Урьд цагт эцэстээ Н авиатай байсан олон үгийн эцсийн Н хаягджээ. Жишээ нь: морин-морь, усун- ус, дарсун-дарс гэх мэт. Ингэж хаягдсан Н авиаг бичихгүй нь мэдээж хэрэг. Гэтэл дээр дурдсан үгсийн эцсийн хаягдсан Н авиа, мөн үгэнд харъяалах,өгөх орших, гарах 3н тийн ялгалын залгаврыг нэмэхэд гарч ирдэг байна. Жишээ нь: Морины, усны, дарсны, моринд, усанд, дарсанд, мориноос, уснаас, дарснаас гэх мэт. Угаасаа тогтворгүй байсан үгийг дээрх 3н тийн ялгалаар хувилгахад тогтворгүй Н гарч ирдэггүй байна. Жишээ нь: ном номын номонд номоос гар гарын гараас гарт гэх мэт. Гэтэл үүнд нь бас гажилт байдаг байна. Тухайн үг нь хоёр өөр утгыг илтгэн бас хоёр өөр үг болж, утгаараа хувирдаг байна. Жишээ нь: өрхийн тэргүүлэгч, өрхний оосор, сумын дарга, сумны зэв, хүүхэд хэлд орох, монгол хэлэнд оруулах, 4дүгээр сарын хоёрнд, 4-ийг хоёрт хуваах гэх мэт [6]. Энэ мэтийн нарийн ялгавар монгол хэлэнд боловсорч байгаа нь эдгээр үгсийг дээрх 3-н нөхцөлөөр цахим дүрмийн аргаар ялгаж, хувиргах боломж одоогоор судлагдаагүй байна. Үүнийг бид энэхүү дүрэмд өөр ерөнхий тэмдэглэгээгээр тусган оруулж шийдвэрлэсэн юм (нЫ, нД, нҮҮс).

Дараагийн өвөрмөц хувиралтай тохиолдол нь хэлний угийн Н үсгээр төгссөн бол Г гийгүүлэгч жийрэглэж, харьяалах, заахын нөхцөл нь эр, эм үг аль ч тохиолдолд ийн, г хэлбэрийн нөхцөлөөр, гарах үйлдэхийн нөхцөлд мөн г гийгүүлэгч жийрэглэж хувирдаг. Энэхүү хувирлын хамгийн гол асуудал нь хэлний үзүүрийн Н төгссөн үгсээс хэлний угийн Н мөн эсэхийг ялгаж таних явдал ба үүнийг дүрмийн аргаар ялгах боломжгүй байсан юм. Иймээс эдгээр үгсийг түүж, толь бичгийн аргаар хувиргахаар шийдвэрлэсэн.

**ҮТ:** с а н 0 + Ы н сан+NNI\_GENI

с а Нг 0 + Ы н

**ХТ:** с а н г и й н сангийн

* + 1. **Дагаврын эгшиг зохицол**

Монгол хэлний дагавар бол нэг үгнээс өөр нэг үг үүсгэх бүтээвэр юм. Тиймээс дагавар бүтээвэр нь цөөн төрөлтэй ба бүх үгэнд орохгүй бөгөөд эгшиг нь мөн тодорхой утгаасаа хамааран хоёроос гурван төрлийн эгшгээр хувирдаг байна. Тэдгээрийг судалж, нэгтгэн дараах байдлаар тэмдэглэн дүрмээ зохиосон болно.

1. Ө : а, и, о

2. И : а, и

3. А : а, э

4. Э : а, о

Дээр өгүүлсэнчлэн цөөн үгэнд хэрэгжүүлэх тул эгшиг зохицох дүрэм гэх мэт бусад дүрмээс арай хялбар юм.

RULE Ө:а => OO [C\*|й C\*] +:@ C\* (Н: (Н:@)) C\* \_\_\_

1. Нэрээс нэр үүсгэх мхӨй, мтгӨй, нхӨй дагавар

**ҮТ:** а й + м х Ө й ай+JVD\_MHAI

**ХТ:** а й 0 м х а й аймхай

1. Нэрээс нэр үүсгэх "гИй"дагавар

**ҮТ:** г э р + г И й гэр+NND\_GAI

**ХТ:** г э р 0 г и й гэргий

1. Нэрээс нэр үүсгэх "дЭй" дагавар

**ҮТ:** о ч + д Э й оч+NND\_DAI

**ХТ:** о ч 0 д о й очдой

1. Нэрээс нэр үүсгэх "CАC "дагавар2

**ҮТ:** э г ч + м А д эгч+NND\_MAD

**ХТ:** э г ч 0 м э д эгчмэд

Энэхүү дөрвөн төрлийн дүрэм нь зөвхөн эдгээр дагаврууд бус түүнтэй төстэй бусад дагаварт мөн адил тэмдэглэгдэх ба өөр дагавруудад бусад эгшиг зохицох тэмдэглэгээг тохируулан тэмдэглэж хувиргасан. (Хавсралт А)

* + 1. **Гээгдэх ба гээгдэхгүй эгшиг**

Зөв бичих дүрмийн дагуу (3.1.2, 3.1.3) дагавар, нөхцөлийг залгахад тухайн үгийн язгуур үндсийн эгшиг гээгдэх, гээгдэхгүй байх, эгшиг жийрэглэх гэх нэлээд өргөн хүрээг хамарсан буюу зохиогдсон бүх хоёр түвшинт дүрмүүдтэй харилцан уялдаатай, судлагдахуун их шаардсан дүрэм. Эгшиг гээгдэхийг устгах үйлдлээр, гээгдэхгүй тохиолдлыг хориглох оператор ашиглаж, эгшиг жийрэглэх тохиолдлыг оруулах үйлдлээр мөн нэмэх тэмдгийн харилцан холбоосуудыг ашиглан зохиосон болно. Жишээ нь: Гээгдэх эгшигтэй

**ҮТ:** г э р э л + Ы г гэрэл+NND\_MAD

**ХТ:** г э р 0 л и й г гэрлийг

Хэлний үзүүрийн Н, хоолойн Г –ийн өмнөх эгшиг гээгдэхгүй

**ҮТ:** с а н а л + Ү Ү р санал+NNI\_INS

**ХТ:** с а н а л 0 а а р саналаар

Эгшиг жийрэглэх

**ҮТ:** ө в г ө д + Д өвгөд+NNI\_DATD

**ХТ:** ө в г ө д ө д өвгөдөд

* + 1. **Е тэмдэглэгээний зохицол**

Энэ нь үйлийн биеэр төгсгөх нөхцөл (я, е, ё)-ийн зөв бичих дүрэм түүний зохицол хамаарахын зэрэгцээгээр тэмдэг үсгийн (ь,ъ,) зохицлыг давхар тооцон зохиогдоно. Тэмдэг үсгийн хоёр түвшинт дүрэм нь тухайн тэмдэг орсон үгэнд боломжит бүтээврүүдээр хувилгахад зөв тохирох ёстой. Хувирсан жишээ:

**ҮТ:** я в + Е яв+VVI\_EP1ASK

**ХТ:** я в ъ я явъя

Тэмдэг үсэг бусад дүрмүүдтэй зохицол

**ҮТ:** б а р ь + У У л барь+NVD\_UUL

**ХТ:** б а р и 0 0 у л бариул

Эдгээр хоёр түвшинт дүрмүүдийг зөв бичгийн дүрмийг ерөнхийд нь баримтлан, хоорондын холбоосуудыг нарийн тодорхойлж, үйлийн ба нэрийн 47-н төрлийн үгийг багтаан, зөв бичих дүрмийн зөрчилгүйгээр нэгтгэсэн нийт 84 дүрэм зохиосон болно.

**4.2 Хоёр түвшинт дүрмийн туршилт**

ПСи-Киммо систем нь дүрмийн файл, үгийн сан гэсэн үндсэн хоёр хэсгийг ашигладаг талаар өмнөх бүлэгт дурдсан билээ. Тэгэхдээ хэлний үндсэн хэсгийг бий болгоход тусгай түлхүүр үгийн тусламжтайгаар файлд тодорхойлон бичиж өгдөг. Дүрмийн файлын бүтэц нь монгол хэлний цагаан толгой, дэд олонлог, дүрмүүд гэсэн үндсэн элементүүдийг зарлахаас эхэлнэ.

*Дүрмийн файл (mon.rgx )-ыг үүсгэх*: Эхний алхамд цагаан толгойгоо үүсгэнэ. ALPHABET гэсэн түлхүүр үгийн дор албан ёсны үгийн сангийн болон өнгөн хэлбэрийн тэмдэгтүүдийг жагсаана. SUBSET дүрэмд хэрэглэх дэд олонлог, PAIRS цагаан толгойгоор үүссэн энгийн харилцан холбоосууд, RULE хоёр түвшинт дүрмийн хэсэг ба энд монгол хэлний үгзүйн 84 дүрмийг байрлуулж ажлуулна..

***Зураг 7.*** *ПСи-Киммо програмын дүрмийн файлын бүтэц*

*Үгийн сангийн файл (mon.lex)- ыг үүсгэх*: Дүрмийн файлыг тодорхойлсноор үгзүйн задлуурыг хийх боломжтой болно. Тухайн хэлний үгийн санг тодорхойлохдоо нэг үндсэн үгийн сан заавал байх шаардлагатай бөгөөд нэр, үйлийн, хавсрал гэх мэт дэд хэд хэдэн үгийн сан байж болно. Эдгээр үгийн санг мөн түлхүүр үгүүдээр зарлаж өгнө.

***Зураг 8.*** *ПСи-Киммо програмын үгийн сангийн файлын бүтэц*

Монгол хэлний үгийн санд нэрийн, үйлийн, бүтээврийн гэсэн гурван дэд олонлогоос бүрдэх ба үгзүйн нэгжүүдийг дэд үгийн сангийн файлд дараах бүтэцтэй хадгална.

Монгол хэлний үгийн сандаа нэрийн 12500, үйлийн 23000 үгийн язгуур үндэс, тодорхойлсон дагаврууд, бүх нөхцөлийг оруулсан. Ингэж боловсруулсан дүрмээ турших үүсгүүр ба задлуурын явцыг схемээр харуулав.

***Зураг 9.*** *Хоёр түвшинт дүрмийн үүсгүүр ба задлуурын схем.*

Хөрвүүлсэн KGEN.EXE бол хоёр түвшинт дүрмийг хүснэгтэн хэлбэрт хувиргана. Жишээ нь: Зөөлний тэмдэг и болохгүй гэсэн дүрмийг хүснэгтэн хэлбэрт хувиргасныг дараах зурагт үзүүлэв.

RULE " ь:и /<= MN \_\_\_ +:0 [II|FK|Ж:]" 4 7

MN ь + Ж FK II @

MN и 0 @ FK II @

1: 2 1 1 1 1 1 1

2: 2 3 1 1 1 1 1

3: 2 1 4 1 1 1 1

4: 2 1 1 0 0 0 1

***Зураг 10.*** *Хоёр түвшинт дүрмийг хүснэгтэн хэлбэрт оруулсан байдал.*

Бид туршилтаа нэрийн, үйлийн бүх төрлүүдийг хамарсан 1000 гаруй үгийг оролтонд өгч хувиргаж үзэхэд 98,4% -тай зөв хувиргаж үүсгэж байгаа ба алдаатай үгсийг авч үзвэл:

1. Зөв бичгийн дүрмийг гаждаг (8) үгс тухайн нэг зарим дүрэмд тохиолдоно. Жишээ нь: сард, өгч, орж, гарч гэх мэт

.

1. Монгол хэлний эгшиг жийрэглэх ба эс жийрэглэх зөв бичгийн дүрэм маш уян хатан, эс жийрэглэх тохиолдлын хоорондын хамаарал ойролцоо байдаг. Мөн эс жийрэглэх тохиолдлыг хориглох оператор ашиглан бичих ба түүнийг хоёр удаа, тусад нь ялгаж бичих боломжгүй байгаагаас дараах хэдэн үгэнд эгшиг жийрэглэхгүй байж чадахгүй байна. Жишээ нь: богинодотугай, дөхөтүгэй

Үүсгүүр, задлуурын гаралтын файл дараах байдалтай байх ба нэг үгний төгсгөлийн хувилбар нь зөв хувилбар болно. Үүсгүүр задлуурын үр дүнг доорх 20 үгэнд хүснэгтээр харуулбал:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Үүсгүүрийн оролт** | **Үүсгүүрийн гаралт** | **Задлуурын оролт** | **Задлуурын гаралт** |
| 1 | ах+Ыг | ахыг | ахыг | ах+Ыг ах+NNI\_OBJ |
| 2 | багаж+Ыг | багажийг | багажийг | багаж+Ыг багаж+NNI\_OBJ |
| 3 | араг+Ыг | аргийг | аргийг | араг+Ыг араг+NNI\_OBJ |
| 4 | зөгий+Ын | зөгийн | зөгийн | зөгий+Ын зөгий+NNI\_GENI |
| 5 | салхи+нЫ | салхины | салхины | салхи+нЫ салхи+CaseGen |
| 6 | өрөө+тЙй | өрөөтэй | өрөөтэй | өрөө+тЙй өрөө+Casecomi |
| 7 | нутаг+Д | нутагт | нутагт | нутаг+Д нутаг+CaseDat |
| 8 | үз+БҮл | үзвэл | үзвэл | үз+БҮл үз+VVI\_CSAPP |
| 9 | зар+лҮгҮ | зарлага | зарлага | зар+лҮгҮ зар+NVD\_LAGA |
| 10 | дулаар+Ж | дулаарч | дулаарч | дулаар+Ж дулаар+VVI\_CPPAR |
| 11 | гар+Е | гаръя | гаръя | гар+Е гар+VVI\_EP1ASK |
| 12 | цохь+Е | цохьё | цохьё | цохь+Е цохь+VVI\_EP1ASK |
| 13 | говь+Ын | говийн | говийн | говь+Ын говь+NNI\_GENI |
| 14 | ор+ҮҮрҮй | ороорой | ороорой | ор+ҮҮрҮй ор+VVI\_EP2ORDR |
| 15 | хар+тУгАй | хартугай | хартугай | хар+тУгАй хар+VVI\_EP3ASKT |
| 16 | нэм+УУзАй | нэмүүзэй | нэмүүзэй | нэм+УУзАй нэм+VVI\_EP2ORDU |
| 17 | хөв+нгУУт | хөвөнгүүт | хөвөнгүүт | хөв+нгУУт хөв+VVI\_CSNGUT |
| 18 | аж+гтУн | ажигтун | ажигтун | аж+гтУн аж+VVI\_EP2ORDG |
| 19 | боловсор+хлҮҮр | боловсрохлоор | боловсрохлоор | боловсор+хлҮҮр боловсор+VVI\_CSFOLL |
| 20 | илэр+мҮгц | илэрмэгц | илэрмэгц | илэр+мҮгц илэр+VVI\_CSPREP |

***Зураг 11****. Үүсгүүр ба задлуурын үр дүнгийн хэсгээс*

**Дүгнэлт**

Цахимд зориулсан судалгаа нь “хүмүүний хэл” –ний зүй тогтол, үйлдэл бүрийг нарийн тодорхойлох шаардлагатай байдаг тул уламжлалт судалгааны онолыг үгзүйн болон хэлзүйн хэлбэр бүр дээр хэрхэн бууж байгааг тооцоолж, түүнийгээ загварчлахыг шаарддаг. Тэрхүү нөхцөлийг хангасан өгөгдөл бүрдсэн тохиолдолд цахим тооцоолуур сайн ажиллана. Тиймээс уламжлалт үгзүйн судалгааны дээр цахимын онцлогийг тусгасан шинэ судалгааг шаарддаг.

Энэхүү судалгааны ажлаар монгол хэл боловсруулалтын системчилсэн хөгжүүлэлтийн эхний шатны гол чиглэл болох цахим үгзүйн боловсруулалт (computational morphology)-ыг хувирах үгнээс хамаарахгүйгээр ПСи-Киммо-ийн хоёр түвшинт загварыг ашиглан монгол хэлний үгзүйн хоёр түвшинт дүрмийг зохиож, туршлаа. Хоёр түвшинт дүрмийг зохиохдоо монгол хэлний зөв бичих дүрэмд тулгуурласан ба “дүрэмд суурилсан” (rule-based) арга хандлагыг хэрэгжүүлсэн болно. Ийм судалгааны ажил урьд өмнө нь нэлээдгүй хийгдэж байсан ч тэр нь хангалттай бус зөвхөн нэрээс нэр үг үүсгэх дагаварын дүрэм гэх мэтийг боловсруулсан бол энхүү судалгааны ажлаар нэр ба үйлийн боломжит бүх дагавар, нөхцөлөөр хувиргадаг болгон сайжруулсан.

Монгол хэлний цахим үгзүйн хүрээнд дагавар ба нөхцөлийг оновчтой ерөнхийлөн тэмдэглэж, залгавар бүтээвэр бүрийг кодын схемийн дагуу кодлон, монгол хэлний хамгийн чухал, өргөн хүрээний зөв бичгийн дүрэм болох эгшиг зохицох ёсыг хоёр түвшинт загвараар дүрэмжүүлэн зөв үүсгэж, задлаж байгаа зэрэг нь шинэлэг тал болсон юм.

Дүрэм зохиох явцад нэрийн 29, үйлийн 18 төрлийн үг бүрд хувирч болох бүх залгавар бүтээврээр зөв хувилах дүрмийг гаргах, тэдгээрийн хоорондын хамаарлын өөр олон дүрмүүд шаардагдах ба тэдгээрийг нэгтгэх, зөрчилгүй болгох нь маш их цаг, ажиглалт, судалгаа шаардсан амаргүй ажил байсан юм. Бусад олон орны энэ аргаар туршигдсан хэлний судалгаанаас монгол хэл нь илүү олон төрлийн үгтэй, хоорондын ялгамж уян хатан шинжтэй, дүрмүүд нь олон бөгөөд ерөнхий нэг дүрэмд захирагдах нь ховор, нэг нь нөгөөгөө үгүйсгэсэн, гажсан хэлбэрүүд ихтэй гэх зэрэг нь хоёр түвшинт дүрмийн өөр илүү шинэ, боломжийг шаардсан ажил байлаа.

Монгол хэлний цахим үгзүйн нийт 84 дүрэм зохиосон ба үгийн санд нэрийн 23000, үйлийн 12500 язгуур үндэс, бүх залгавар бүтээвэр оруулан тэдгээрийг нэр ба үйлийн нийт 47-н төрлийн үгний төрөл бүрээс 4, 4-н (а, э, о, ө орсон) үгийг сонгон авч бүх нөхцөл боломжит дагавраар хувиргах 1000 гаруй үгэнд туршиж үзэхэд 98,4 хувьтай боловсруулж байна. Эндээс алдаа гарсан үгнүүдийг авч үзэхэд:

1. Зөв бичгийн дүрмийг гаждаг үгс.(8) Жишээ нь: сард, өгч, орж, гарч гэх мэт. Энэхүү цөөн үгийг толь бичгийн аргаар шийдэх боломжтой.
2. Эгшиг эс жийрэглэх тохиолдлыг хориглох оператор ашиглан бичих ба түүнийг хоёр түвшинт загварт хоёр удаа, тусад нь ялгаж бичих боломжгүй нь Д төгссөн хэдэн үгэнд эгшиг жийрэглэж буруу гарч байна. Үүнийг монгол хэлний үгзүйн эгшиг жийрэглэх дүрэм, хоёр түвшинт дүрэм хоорондын уялдааг тусгайлан илүү нарийн боловсруулах шаардлагатай.

**Цаашдын судалгаа:**

Дээрх дүрмүүд нь залгаврыг нэг нэгээр оролтод өгөхөд гүйцэд хувиргаж байгаа ч харин нэгэн зэрэг дараалан орсон залгаврыг үүсгэх задлахад бүрэн үүсгэж, задлаж чадахгүй байгаа юм. Үүнийг сайжруулан боловсруулах хэрэгтэй.

Цаашид монгол хэлний үгзүйн зөв бичих дүрмийг илүү нарийвчлан судалж, сайжруулах, тухайлбал боловсруултад бэрхшээл учруулж буй монгол хэлний тогтворгүй Н авианы дүрмийг цэгцлэх шаардлагатай. Ийм төрлийн нөхцөлийг өөр ерөнхий тэмдэглэгээг (нЫ, нД, нҮҮс) нэмж оруулан шийдвэрлэсэн. Гэсэн хэдий ч энэ нь хувирах үгнээс хамаарч байгаа тул цахим үгзүйд дутагдалтай юм. Тиймээс гүйцэд нэгэн тийш болгоход ихээхэн нарийн судалгаа шаардлагатай юм гэдгийг ойлгосон. Жишээлбэл : ууланд гарах, уулнаас буух гэж хэлээд уулны өвөрт гэхгүй уулын өвөрт гэж хувилах юм. Бидний дээр хэлсэн гурван тийн ялгалаар Н жийрэглэсэн нөхцөлийг авна гэсэн дүгнэлт, дүрэмд захирагдахгүй зөрчилдөж байгаа юм. Энэ асуудлыг хэл шинжлэлийн судлаач нар цаашдаа сайн судалж байж шийдвэрлэх хэрэгтэй юм.

**Ном зүй**

[1] Г. Амарсанаа, МУИС, 2005 “Монгол хэлний тэмдэглүүр”

[2] Ж.Пүрэв, Ч.Алтангэрэл, МУИС, 2003 “Монгол хэлний үсэг, бичгийн талаарх судалгаа” гар бичмэл

[3] Ц. Өнөрбаян, 2006 “Орчин цагийн монгол хэлний үгзүй ”

[4] Ж.Пүрэв, Г.Амарсанаа, МУИС, 2007 “Юникод ПСи-Киммо: Олон хэлэнд зориулсан үгийн шинжлүүр”

[5] Коскенними “A disco**very** procedure for two level morphology” өгүүлэл

[6] Ц.Дамдинсүрэн, Б.Осор, 1983 “Монгол үсгийн дүрмийн толь”

[7] S.M. Shieber, “An introduction to unification based approaches to grammar”, CSLI Lecture Notes No. 4. Stanford, CA, 1986.

[8] E.L. Antworth. “Morphological Parsing with Unification-based Word Grammar”, a paper presented at North Texas Natural Language Processing Workshop, May 23, 1994

[9] URL: http://www.sil.org/pckimmo/, 2007.11

[10] Kemal oflazer, Elvan Gocmen,Cem Borzsahin “An Outline of Turkish Morphology” 1994

[11] “Englex: A Computational morphology of English” 1995

[12] Ш.Чоймаа, МУИС, 2010 “Хөрвүүлэх програмд зориулсан монгол хэлний судалгаа”

[13] Ж.Пүрэв, Ч.Алтангэрэл, З.Цолмон, 2005 “PC-KIMMO-based Description of Mongolian Morphology”

[14] Ч.Одбаяр, МУИС, 2008 “Монгол хэлний үгзүйн суурь загварчлал”

[15] Г. Түмэнбаяр, МУИС, 2006 “Хоёр түвшинт үгзүйн загварыг монгол хэлний хэл шинжлэлд хэрэглэх”

[16] Kemal oflazer, “Two level Description of Turkish Morphology” 1995

**Хавсралт А**: Нэрийн дагавар, нөхцөлийн ерөнхий тэмдэгээ, кодын хүснэгтийн хэсгээс

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Affixes | Morphemic Representation | | | Code | Gloss | | | Example |
| ууд, үүд | УУд | | | NNI \_PLUU | Plural УУд | | | салбар-ууд. мөр-үүд |
| чууд, чүүд | чУУд | | | NNI \_PLUG | Plural чУУд | | | залуу-чууддөгш-чүүд |
| нар | нар | | | NNI \_PLUN | Plural нар | | | эмч нар |
| Д | д | | | NNI \_PLUD | Plural д | | | сурагчид |
| с | с | | | NNI \_PLUS | Plural c | | | залуус |
| 0 |  | | | NNI\_NOM | Nominative Case | | | бал,гэр |
| ийн,ын, ий, ы, н | Ын | | | NNI \_GENI | Genitive Case | | | балын,гэрийн, нохойн,далайн |
| д, т | Д | | | NNI \_DATD | Dative Case | | | ахад,эгчид,дүүд,цаст |
| аа,ээ,оо,өө | ҮҮ | | | NNI \_DATY | Dative Case | | | уулнаа,хотноо,өргөөнөө |
| ыг, г, ийг | Ыг | | | NNI \_OBJY | Objective Case | | | ахыг.эгчийг, нохойг.дүүг,санг |
| аас,ээс,оос,өөс | ҮҮс | | | NNI \_ABL | Ablative Case | | | ахаас,эгчээс,дүүгээс,модноос |
| аар,ээр,оор,өөр | (г)ҮҮр | | | NNI\_INS | Instrumental Case | | | ахаар,эгчээр,дүүгээр.нохойго ор,өрөөгөөр |
| тай,тэй,той | тЙй | | | NNI\_COMT | Comitative Case | | | ахтай . эгчтэй ,д үүтэй , нохойтой ,өрөөтөй |
| РУУ,РҮҮ | рУУ | | | NNI \_COMR | Directive Case | | | ах руу,эгч рүү |
| луу,лүү | | лУУ | NNI \_COML | | | Directive Case | нэр лүү,сар луу | |
| аад,ээд,оод;өөд | | ҮҮд | NNI \_СОМҮ. | | | Directive Case | уул өөд | |
| аа,ээ,оо,өө | | (г)ҮҮ | NNI \_POSSY | | | General Poss. | ахаа,эгчээ,дүүгээ | |
| минь | | минь | NNI \_POSS1I | | | 1st pers. Poss. | ах минь,эгч минь | |
| маань | | маань | NNI \_POSS1 | | | 1st pers. Poss. | дүү маань,ах маань | |
| чинь | | чинь | NNI \_POSS2I | | | 2nd pers. poss. | ах чинь,эгч чинь | |
| тань | | тань | NNI \_POSS2I | | | 2nd pers. poss. | дүү тань,ах тань | |
| нь | | нь | NNI \_POSS3 | | | 3rd pers. poss. | ах ньдүү нь | |
|  | |  |  | | |  |  | |

1.2 Нэр үг үүсгэх дагаврууд (NxDxxxxx)

1.2.1 Нэрээс нэр үүсгэх дагавар (NND\_\_xxxxxx)

|  |  |  |  |
| --- | --- | --- | --- |
| Affixes | Morphemic Representation | Code | Example |
| ra, гэ, го, гө | гҮ | NND \_GA | гадар-ra, тоос-го |
| гана, гэнэ, гоно, гөнө | гҮнҮ | NND \_GANA | хар-гана, бөөр-гөнө |
| лаг, лэг. лог, лөг | лҮг | NND \_LAG | ур-лаг, үдэш-лэг, шор-лог |
| лан, лэн, лон, лөн | лҮн | NND \_LAN | тариа-лан, шүд-лэн, соёо-лон, өнгө-лөн |
| тан, тэн, тон, төн | тҮн | NND \_TAN | оюу-тан, бүс-тэн, ном-тон, өс-төн |
| таг, тэг, тог, төг | тҮг | NND \_TAG | бур-таг, өр-төг |
| цаг, цэг, цог, цөг | цҮг | NND \_CAG | эрэг-цэг, дов-цог |
| нцаг, нцэг, нцог, нцөг | нцҮг | NND\_NCAG | зээр-(э)-нцэг, оро-(о)-нцог |
| нцар, нцэр, нцор, нцөр | (Ү)нцҮр | NND \_NCAR | гуч-(и)-нцар, зээ-нцэр |
| гтай, гтэй, гтой | (Ү)гтИй | NND\_GTAI | хат-(а)-гтай. эмэ-(э)-гтэй |
| лдай. лдэй, лдой | (Ү)лдЙй | NND \_LDAI | авга-лдай, хүүх-(э)-лдэй, бор-(о)-лдой |
| дай, дой | дОй | NND \_DAI | бухан-дай, оч-дой |
| ган, гэн | гАн | NND \_GAN | бэр-гэн, эм-гэн |
| маг, мэг | мАг | NND \_MAG | зайр-маг, ир-мэг |
| мад, мэд | мАд | NND \_MAD | ах-мад, эгч-мэд |
| cap, сэр | сАр | NND \_SAR | ам-сар, зав-сар |
| гай, гий | гИй | NND \_GAI | ам-гаи, гэр-гии |
| ж | ж | NND \_J | гун-ж |
| жин | жин | NND \_JIN | алт-жин |
| лж | лж | NND \_LJ | тооно-лж |
| мж | мж | NND \_MJ | арга-мж |
| бай | бай | NND \_BAI | тал-бай |
| баа | баа | NND \_BAA | ул-баа |
| бар | бар | NND \_BAR | тал-бар |
| ч | ч | NND \_CH | эм-ч |
| чин | чин | NND \_CHIN | зураг-чин |
| л | л | NND \_L | үеэ-л |
| ль | ль | NND \_LI | хаяа-ль |
| вч | вч | NND \_VCH | хүзүү-вч |
| д | ц | NND \_C | бор-ц, өр-ц |
| с | с | NND \_S | дүр-с, гөрөө-с, үйл-с |
| 3 | 3 | NND \_Z | хар-з, ар-з, хор-з, шар-з |
| т | т | NND \_T | ам-т, шим-т |
| н | (Р)н | NND \_N | од-(о)-н, шор-(о)-н |
| мт | (Р)мт | NND \_MT | гол-(о)-мт |
| вх | (Р)вх | NND \_VX | ид-(э)-вх |
| цээ | цээ | NND \_CEE | хэрэг-цээ |
| X | X | NND \_X | дээр-х, доор-х |
| дахь, дэх, дох, дөх | дҮх(ь) | NND \_DAXI | уул-(ан)-дахь, өрөө-(н)-дөх, |
| ийнх, ынх | Ынх | NND \_IINX | ах-ынх, эгч-ийнх |
| ийнхэн, ийнхөн, ынхон, | ЫнхҮн | NND \_IINXEN | гэр-ийнхэн, төр-ийнхөн,хот-ынхон, |

|  |  |  |  |
| --- | --- | --- | --- |
| нх | нх | XNT\_LXX | хангай-нх |
| нхан, нхэн, нхон, нхөн | нхҮн | XXD\_NXAN | хангай-нхан, толгой-нхон, хэрээ-нхэн, өргөө-нхөн |

1.2.2 Үйлээс нэр үг үүсгэх дагавар (NVD\_xxxxxx)

|  |  |  |  |
| --- | --- | --- | --- |
| Affixes | Morphemic Code Representation | | Example |
| aac, ээс, оос, өөс | ҮҮс | NVD\_AAS | зур-аас, хэр-ээс, огто(о->0)л-оос, нөх-өөс |
| aap, ээр, оор, өөр | ҮҮр | NVD \_AAR | хаз-аар, үд-ээр |
| аач, ээч, ооч, өөч | ҮҮч | NVD\_AACH | зур-аач, бич-ээч, номло-(о->0)-ооч, гөрд-өөч |
| бар. бэр, бор. бөр вар, вэр, вор, вөр | БҮр | NVD\_BAR | тайл-бар, сийл-бэр, цоол-бор, хөтөл-бөр, дай-вар, үз-вэр, олд-вор |
| га, гэ, го, гө | гҮ | NVD \_GA | оруул-ra, тул-га. дар-га |
| дал, дэл, дол, дөл | дҮл | NVD \_DAL | буу-дал, орго-дол, өргө-дөл |
| дас, дэс, дос. дөс | дҮс, | NVD\_ DAS | цуул~дас, хив-дэс |
| лаг, лэг, лог, лөг | лҮг | NVD \_LAG | өр-лөг, үз-лэг, бар-лаг |
| лан, лэн, лон, лөн | лҮн | NVD \_LAN | хад-лан, зов-лон, өлсгө-лөн |
| лага, лого | лҮгҮ | NVD \_LAGA | зар-лага, ор-лого |
| лга, лго | лгҮ | NVD \_LGA | барь-(ь->и)-лга, зорь-(ь->и)-лго |
| маг, мэг, мог, мөг | мҮг | NVD \_MAG | зуур-маг, ээд-мэг, ороо-мог |