

Монгол яриаг изохрон шинжээр хэсэгчилэх нь

Praat програм дээр автоматжуулах

Төрийн Ууганбаяр¹, Чагнаагийн Алтангэрэл²

Машин оюуны лаборатори, Мэдээлэл, компьютерийн ухааны тэнхим
Хэрэглээний Шинжлэх Ухаан Инженерчлэлийн Сургууль
Монгол Улсын Их Сургууль
Улаанбаатар, Монгол Улс

uugaa_t@yahoo.com¹, altangerel@seas.num.edu.mn²

Хураангуй — Энэ судалгааны ажлаар Монгол яриаг изохрон шинжээр автоматаар хэсэгчилэн яриа таних системийн хэсэг болох корпусыг үүсгэх ажлыг гүйцэтгэв.

Түлхүүр үг — Монгол хэл, яриа таних, изохрон, корпус

I. УДИРТГАЛ

A. Яриа таних систем

Орчин үед өгөгдөл, мэдээллийг хүн ба машин хооронд дамжуулахад хялбар болгох үүднээс хэл яриаг автоматаар танин текст хэлбэрт хөрвүүлэх, текстийг хэл ярианд хөрвүүлэх системүүд боловсронгуй болсоор байна. Хамгийн энгийн жишээний нэг та бидний өдөр тутам хэрэглэдэг ухаалаг төхөөрөмжүүдэд хэл ярианы модулууд дагалдан ирдэг болсон. Уг системүүдэд дэлхийн түгээмэл хэрэглэгддэг ихэнх хэлүүд орсон ба Монгол хэлний ярианаас текст болгох систем одоогоор хөгжөөгүй байна. Иймээс Монгол хэлний яриа таних систем хөгжүүлэхэд шаардлагатай нэг хэсэг болох ярианы корпус бий болгоход энэ судалгааны ажлын нэг чиглэл болгож авч үзлээ.

B. Яриа таних арга

Анх 1932 оноос АНУ-ын Bell Labs-д тооцоолох электрон машин үүсэх үеэс хүний үг, яриаг таниж өгөгдөл болон команд болгон хувиргах санаа үүсжээ. Түүнээс хойш маш олон төрлийн арга, техник үүсэн хөгжиж байна. Эдгээрээс хамгийн түгээмэл арга болох Нуугдмал Марковын Загвар (НММ) нь статистик дээр суурилан ажилладаг бөгөөд одоо үед ихэнх системүүд тус арга дээр суурилан хөгжиж байна.

Статистик дээр үндэслэсэн яриа таних аргуудад сургалтын өгөгдөл (training data) хэрэглэгддэг бөгөөд яриа таних систем болгон корпус гэж нэрлэгдэх ийм өгөгдлийн сантай байдаг. Танилтын хувь өндөр системүүд корпустай хэдэн зуун сая нэгжтэй байдаг. Монгол Улсын Их Сургуулийн машин оюуны лаборатори дээрх Монгол хэлний корпус одоогоор цөөн тооны нэгжтэй байгаа бөгөөд уг корпусыг өргөжүүлэх ажил удаашралтай явагдаж байна. Хамгийн том бэрхшээл бол корпус үүсгэхэд гар ажиллагаа ихээр шаардлагатайгаас цаг хугацаа ихээр зарцуулдаг болно.

Европын болон Японы их сургуулиудад яриа, текстийг автоматаар хэсэгчилэн, корпус бүрдүүлдэг боловсронгуй олон системүүд ашиглагдаж байна.

II. ТАНИЛЦУУЛГА

A. Судалгаанд хэрэглэсэн өгөгдөл

Тус судалгаанд бид Нацагдоржийн номын сан дээр байрлах “Daisy” ярьдаг ном бүтээх төслийн хүрээнд Ц. Оюунгэрэлийн “Америкт суралцсан тэмдэглэл” номны дууны өгөгдлийг ашиглав. Тус номны нийт 25 файл бүхий 2,85 ГБ хэлбэрийн дууны өгөгдөл (wav), харгалзах текст өгөгдлийг ocr програм ашиглан оруулсан болно. Текст өгөгдөлд номны эхний 3 бүлгийн 170 хуудас материалаас тоо, гадаад үг, тусгай тэмдэглэгээ зэргийг хөрвүүлсэн зөвхөн крилл үсгээр бичигдсэн эх болгосон болно. Өгүүлбэр хоорондын заагийг цэгээр, үг хоорондын заагийг хоосон зай, үе, авиа хоорондын зайг компьютер хэл шинжлэлийн төвийн syllable¹ програмаар ялган тэмдэглэв.

B. Praat

Praat гэдэг нь Голланд хэлнээс гаралтай ярих гэсэн үг бөгөөд Амстердамын их сургуулийн хэл шинжлэлийн профессор Paul Boersma, David Weenink нар анх 1995 онд боловсруулан гаргасан яриа болон дууны мэдээлэл боловсруулах хүчирхэг хэрэгсэл юм. Одоо үед уг системийг олон судлаач, эрдэмтэд, компаниуд хэрэглэж байна. Уг систем нь өөрийн script хэлтэй бөгөөд үүгээр яриа боловсруулах том өгөгдөл, мэдээллийг түргэн хугацаанд, автоматаар боловсруулах давуу талтай болгодог. Та бүхэн www.praat.org хаягаар тус системийг хэрэглэх боломжтой.

B. Ярианы изохрон чанар

Ярианы хэлэнд өргөлтөд болон үет хэлүүдэд авианы үеүд ижил хугацаанд хэлэгддэгийг анх английн авиа зүйч А. Lloyd James 1940 онд таамаглажээ. Энэ шинжийг дэлхийн түгээмэл олон хэлүүдэд мөн эртний хэлнүүдэд хүртэл хангаж байгааг ажиглан тогтоосон байна. Хэл шинжээч, мэдрэлийн эмч Eric Heinz

¹<http://milab.num.edu.mn> Lenneberg өөрийн “Хэл яриан биологийн үндэс” (1967) номондоо ярианы хэмнэл ойролцоогоор 160 ± 20 миллсекунд гэж тогтоосон байна. Үүгээр ярианы хэмнэлийг авиазүй, хэлзүйн

хувьд бус физиологийн хувьд тогтмол байна гэж үзсэн байна.

Хүний яриа амьсгал хоорондоо өгүүлбэр, үг, үе, үсгийг ижил буюу ойролцоо хугацаанд хуваарилагдан хэлдэг. Энэ нь тухайн хүний нас, ярианы хурд, шууд яриа, харилцах яриа, уншиж ярих зэргээс хамаарна. Бидний тохиолдолд манай өгөгдлийг уншсан хүн эх зохиолыг харж, харьцангуй тайвнаар, тогтмол хурдаар уншсан байв.

III. ӨГҮҮЛБЭР ХООРОНД ЗААГЛАХ

Өгүүлбэр хооронд зааглах нь изинхрон шинжтэй хамаарал багатай бөгөөд урт өгүүлбэрийг уншигч урт хугацаанд, богино өгүүлбэрийг богино хугацаанд унших нь тодорхой юм. Өгүүлбэр хооронд зааглах нь нийт ярианы өгөгдлөөс чимээгүй үе (silence)–ийг тогтоох, дараагийн шатны хэсэгчлэл болох үг хооронд зааглах ажиллагаанд урьдчилан боловсруулалт хийхэд дөхөм болно.

Бидний судалгаанд текст өгөгдлүүдийг дахин боловсруулж өгүүлбэр хэлбэрт оруулсан. MS Word програмаар өгүүлбэрийн дунд ямар ч тусгай тэмдэг үлдээлгүйгээр зөвхөн өгүүлбэрийн төгсгөлд тавьдаг цэгээр өгүүлбэр бүрийг бусдаас бүрэн ялгав. Нийт 2462 өгүүлбэр уншигдсан ба дундаж өгүүлбэр 15,16 тооны үгтэй байв.

Зураг 1. Өгүүлбэр, чимээгүй үеийн зааглалт



Зураг дээрээс praat програм дээр дууны өгөгдлийг (sound) текст өгөгдөлтэй (TextGrid) харьцуулан харуулсан байна. Чимээгүй үеийг sil (silent) тэмдгээр харуулсан.

Дууны өгөгдөлд уншигч харьцангуй тогтмол хурдаар уншсан тул зөвхөн чимээгүй үеийг олж тогтоох хялбар байв. Чимээгүй үеийг бид дууны өгөгдлөөс эрчимийн шинжилгээ (Intensity analysis) хийж үндсэн давтамжийн (Pitch) муруйтай харьцуулан тодорхойлсон болно.

Автоматжуулах үүднээс чимээгүй үеийг олох, текст өгөгдлөөс өгүүлбэр бүрийг ярианы үед харгалзуулан оруулах Praat script бичиж гүйцэтгэлээ. Текст өгөгдлийг Praat програмын TextGrid өргөтгөлтэй хугацаа, текст агуулсан файлд хадгалсан болно.

Praat-аар нийт 2387 өгүүлбэр танигдсан бөгөөд энэ нь номын өгүүлбэрээс 75-аар их буюу 96.95%-ийн алдаатай байв. Энэ нь уншигч бүлгийн нэр, гарчиг, зарим тодорхойлолт, зохиогчийн бусадтай ярилцсан яриа зэрэгт зогсоц авч байсан ба зарим урт өгүүлбэрийн дунд амьсгал авсан, дараалсан хоёр

богино өгүүлбэрийг шууд хэлсэн зэргээс дууны болон текстэн өгөгдөлүүд зөрүүтэй гарсан байв.

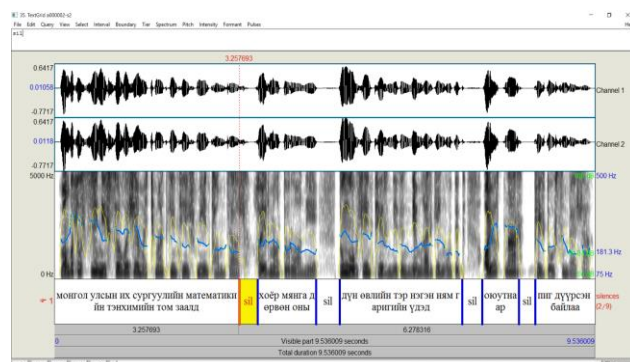
IV. ҮГИЙН ИЗОХРОН ШИНЖ

Өгүүлбэрт изохрон шинж ажиглагдахааргүй бол үг хэлэхэд яригч этгээд өгүүлбэр хэлсэн хугацаанд тогтмол хурдтай буюу үгийн уртаас үл хамааран ижил хугацаанд хэлэх хандлагатай байдаг. Энэ нь богино үгийг сунган, урт үгийг шахаж дундаж хугацаанд хэлдэг. Үргэлжилсэн яриа, уншсан ярианд хүн уг үйлдлийг анзааралгүйгээр, автоматаар хийдэг.

Өгүүлбэр дотор үг хэлсэн хугацааг тогтооход чимээгүй үеийг мөн тодорхойлох шаардлагатай болно. Энэ чимээгүй үе нь өгүүлбэр хоорондын чимээгүй үеээс харьцангуй бага хугацаанд үргэлжилдэг. Өгүүлбэр дотор яригч амьсгал авах тохиолдол ховор бөгөөд өгүүлбэрээ төгсгөж байж амьсгал авдаг. Харин уншигч текстийг харж уншсан бөгөөд зарим өгүүлбэр, үг дараагийн мөрнөөс үргэлжлэх тул мөр шилжих үед чимээгүй үе үргэлжлэх боломжтой.

Чимээгүй үеийг Praat –ийн өөрийн функц ашиглан хамгийн бага чимээгүй үеийн 0,2 секунд гэж тооцсон. Зарим тохиолдолд тэсрэлтэг гийгүүлэгчийн нүргээн үргэлжлэх хугацаа, урт эгшиг хэлэх үед дараагийн гийгүүлэгчийн эхлэх хугацааг чимээгүй үе гэж тодорхойлсон байв.

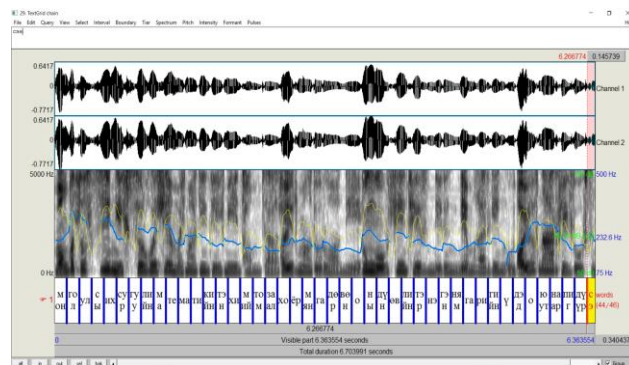
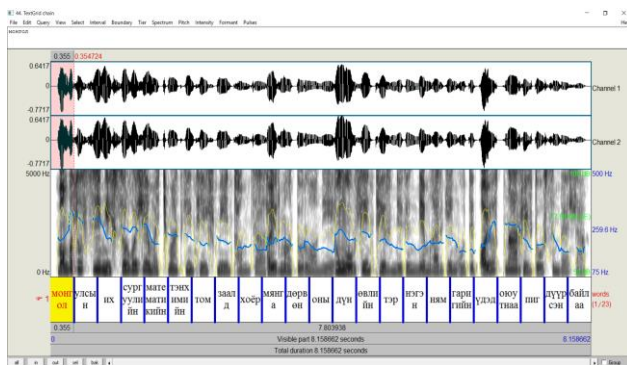
Зураг 2. Өгүүлбэр доторх чимээгүй үеийг тооцсон байдал



Чимээгүй үеийг дууны өгөгдлөөс хасаж, үлдсэн хэсгийг тухайн өгүүлбэрт хэлсэн үгийн тоонд ижил хугацаагаар хуваан үзэснийг Зураг 3-аас харж болно.

Нийт 37342 үг уншигдсан ба үүнд ч, л зэрэг үсэг дангаараа сул үгэнд тооцогдож байхад хамгийн урт үг “интернационалчлагдаж” 20 үсэгтэй байв. Дундаж үг 5,77 үсэгтэй бөгөөд үгийн уншигдах байдал үгийн уртаас хамаарч байв. Дунджаар 6 хүртэл үсэгтэй үгийг ойролцоогоор 0,35 секундэд, 7 ба түүнээс дээш үсэгтэй үгэнд 0,37-оос 0,49 секунд хүртэл хугацаанд хэлж байсан болно. Зарим харьцангуй урт үгнүүд, өгүүлбэрийн гол утга гаргах үг, үйл үгнүүд бусдаас урт хэлэгдэж байв.

Зураг 3. Үгийн тоонд ижил хугацаагаар хуваах



Амьсгалын хэмнэлээс шалтгаалан өгүүлбэрийн эхэнд байгаа үг чанга, урт хэлэгдэж байхад өгүүлбэрийн сүүлийн үгнүүд харьцангуй богино хугацаанд хэлэгдэх хандлагатай байв. Бүх тооны нэр, тэмдэг нэрнүүд хоорондоо адил хугацаанд хэлэгдсэн болно.

Текстэн өгөдлөөс хамгийн их давтагдсан үгнүүд бол “нь” 670 удаа, “гэж” 346 удаа, “ч” 330 удаа, “нэг” 273 удаа, “би” 268 удаа, “юм” 265 удаа, “их” 264 удаа, “байна” 256 удаа, “энэ” 236 удаа, “байгаа” 256 удаа тус тус хэлэгдсэн байв.

Дууны өгөгдлийг ямар нэгэн шинжээр ангилахгүйгээр шууд хэсэгчилэн хувааж байгаа тул алдаа гарах магадлал өндөр бөгөөд алдсан хэсгээс цааших бүх хэсгүүд буруу байна гэдэг нь тодорхой.

V. ҮЕИЙН ИЗОХРОН ШИНЖ

Хэлний изохрон шинж үеийн түвшинд харьцангуй сайн ажиглагдана. Монгол хэлэнд эгшиг гийгүүлэгч сөөлжилт их, олон гийгүүлэгч давхардан орох тохиолдол бага байдаг бөгөөд “эгш”, “эгш+гий”, “гий+эгш”, “гий+эгш+гий”, “гий+эгш+эгш+гий” “гий+эгш+гий+гий” дарааллаар үе үүсдэг. Ихэнд үгнүүд 2-3 үенээс бүрдэг бөгөөд нэг үетэй үг цөөхөн байдаг.

Үг доторх чимээгүй үе зарим 0,02 секундын үргэлжлэх хүрээнд ажиглагдаж байсан. Үндсэн давтамжийн тасралтын цэгүүдэд буюу эгшиг хэлэгддэг хэсэгт хэрэв эрчим бага хэмжээтэй болсон бол чимээгүй үед тооцогдохоор байв.

Чимээгүй үеийг хасаж, үгийн текстийг syllable програмаар үе болгон задалж дууны өгөгдөл дээр шууд үеийн тоогоор хувааж харуулбал Зураг 4 гарна.

Бидний шинжилгээгээр нийт 77246 үе бүртгэгдсэн ба хамгийн урт үеүүд бол “гууль”, “хуульч” хоёр байсан бөгөөд хамгийн богино үеүүд бол дангаар орсон эгшиг, зарим тохиолдолд гийгүүлэгчүүд байв. Үеийн дундаж урт 2,78 үсэгээс бүтсэн байв.

Үгийн дотор үеийн тоонд ижил хэмжээгээр хуваасаныг Зураг 4 дээр харуулсан байна. Үеийн уртаас хамаарах хамааралаар богино үе (1-3 үсэгт), урт үе (4-6 үсэгт) гэж ангилан үзэж болхоор байв.

Зураг 4. Үеийн тоонд ижил хугацаагаар хуваах

Ихэнх үеүүд хэрэв өмнөх үеүүдэд алдаа гараагүй бол зөв хэсэгт хуваагдсан байсан. Нэг чухал нөлөөлөл нь шуугиант гийгүүлэгчид уртаар хэлэгдэх үед, задгай үеийн эгшигийн үргэлжлэл дараагийн үеүүдэд зарим хэсэг нь дуудагдаж байв. Нэг үсгээс бүтсэн үе, 4 ба 5 үсгээс бүтсэн үеээс 0,05 секунд богино хугацаанд дуудагдаж байлаа.

Үе ба түүнээс доош түвшинд изохрон чанар харьцангуй өгсөх хандлагатай ч дууг сонсох үед сонсогч этгээдийн туршлагаас хамаарах магадлалтай. Дуудагдах үеүүд зөв хувиарлагдаж байгааг хүлээн авах тест (reception test)-ээр тодорхой тооны туршилтын олонлог дээр гүйцэтгэж тогтоох боломжтой юм.

VI. АВИАНЫ ИЗОХРОН ШИНЖ

Хэл шинжлэлийн хувьд авиа зүйн түвшинд утга ялгах хамгийн бага хэсгийг авиа (phon) гэж нэрлэнэ. Эгшигээс гарах авиа өгүүлэхүйн эрхтэний оролцоогүйгээр чөлөөтэй урсан гардаг тул дуу, гийгүүлэгчээс гарах авиа ямарваа нэгэн саадыг давж гардаг тул анир тус тус гаргана.

Авианы өмнө үеийн түвшинд чимээгүй үеийг арилгасан тул дууны өгөгдлийг авианы тэмдэглэгээтэй харьцуулан шууд хуваасан болно. Үеийг авианд задлаж, авиаг тусгай тэмдэглэгээгээр тэмдэглэх ажлыг (grapheme to phoneme conversion) syllable програм дээр гүйцэтгэв. Жишээ нь: Монгол гэдэг үгийг мон-гол гэсэн хоёр үед цааш нь m-o1-ng-G-o1-1 гэсэн зургаан авианд задарч байна.

Бидний ашигласан текстэн өгөдлийг авиа зүйн түвшин хүртэл задалбал нийт 57 төрлийн 388662 авиа тоологдсон болно. Хамгийн их давтагдсан авиа бол “a1” 16101, “g” 13864 удаа, “l” 12813 удаа, “e1” 11554 удаа, “i” 11444 удаа, “ng” 10563 удаа тус тус давтагдсан байв.

Бидний хуваалт авиа тус бүрийн онцлог үл харгалзан шууд хуваасан бөгөөд сонсгол зүйн тест хийж үзсэний үндсэн дээр зөв хуваасан эсэхийг тодорхойлох боломжтой байв.

VII. ЯРИАНЫ ХУРД

Манай тохиолдолд яригчийн ярианы дундаж хурдыг нэг секундэд хэдэн өгүүлбэр, үг, үе, авиа хэлж байгааг тооцолсоныг Хүснэгт 1-т харуулав.

Хүснэгт 1 Яригчийн ярианы дундаж хурд

Хэсэгчилсэн хүрээ	Яригчийн дундаж хурд
Өгүүлбэр	0,03 өг/сек
Үг	0,42 үг/сек
Үе	1,26 үе/сек
Авиа	4,48 ав/сек

VIII. ДҮГНЭЛТ

Монгол хэл нь бусад хэлний адил изохрон шинж агуулж байгаа эсэх, цаашид хамгийн бага түвшинд корпус үүсгэх боломжтой эсэхийг судалгаагаар тогтоохыг зорилоо.

Изохрон шинжийг тогтоохын урьд чимээгүй үе гэж нэрлэгдэх яригч этгээдийн зогсоц, зарим төрлийн анир зэргийг урьдчилан тооцож, дууны хэсгээс автоматаар хассан. Тус үйлдлийг гүйцэтгээгүй тохиолдолд изохрон хуваалтанд алдаа гарах магадлал өндөр байсан.

Үгийн түвшин изохрон шинж 2-3 үетэй, 6 хүртэл үсэгтэй үгэнд бүрэн ажиглагдаж байхад үеийн түвшинд бүтэн үеүүдэд яригч изохрон байж чадаж байв. Авиа зүйн түвшинд бид зөв хуваасан эсэхийг сонсголын тестийн тусламжтай тогтоох боломжтой юм.

Бүрэн зөв танигдсан үе, авианы дууны өгөгдлийг тусад нь хадгалан корпус үүсгэх боломжтой бөгөөд цаашид статистик шинжилгээнд оруулан Марковын

коэффициент гарган авч, яриа таних системд сургалтын өгөгдлөөр ашиглах боломжтой.

АШИГЛАСАН МАТЕРИАЛ

- [1] <http://www.fon.hum.uva.nl/praat/manual/>
- [2] Perceptual isochrony and fluency in speech by normal talkers under varying task demands, Laura C. Dilley, Jessica Wallace & Christopher C. Heffner, Lansing, USA
- [3] Time discrimination in a monotonic, isochronous sequence, Anders Friberg and Johan Sundberg Royal Institute of Technology, Department of Speech Communication and Music Acoustics, Box 700 14, S-100 44 Stockholm, Sweden
- [4] EasyAlign: an automatic phonetic alignment tool under Praat, Jean-Philippe Goldman Department of Linguistics, University of Geneva, Switzerland,
- [5] Orthographic Transcription: which Enrichment is required for phonetization?, Brigitte Bigi, Pauline Péri, Roxane Bertrand, Laboratoire Parole et Langage CNRS & Aix-Marseille Université
- [6] SegProso: A Praat-Based Tool for the Automatic Detection and Annotation of Prosodic Boundaries in Speech Corpora, Juan María Garrido, Department of Translation and Language Sciences, Pompeu Fabra University
- [7] SPPAS: a tool for the phonetic segmentation of speech, Brigitte Bigi, Laboratoire Parole et Langage CNRS & Aix-Marseille Université
- [8] Praat script to detect syllable nuclei and measure speech rate automatically, Nivja H. de Jong, Ton Wempe, University of Amsterdam
- [9] Automatic Speech Segmentation for Italian (ASSI) Tools, Models, Evaluation and Applications, Francesco Gangemi, Francesco Cutugno, Bogdan Ludusan, Dino Seppi, Dirk Van Compernelle