

UDK: 004.912:004.738.4

TABIIY TILNI QAYTA ISHLASH YORDAMIDA NUTQNI IMO-ISHORA TILIGA TARJIMA QILISH TEXNOLOGIYALARI

Jurayev Dilshod Boymuradovich,

Tayanch doktoranti

dilsamtuit@gmail.com

TATU

Anotatsiya. Maskur maqolada O'zbekistondagi eshitish va gapirish qobiliyati zaif insonlar uchun nutq va matndan Surdo imo-ishora tiliga tarjima tizimini yaratish g'oyasi taqdim etilgan. Tizim nutq va matnni qabul qilib, tabiiy tilni qayta ishlash moduli yordamida tahlil qiladi hamda ma'lumotlar bazasidan mos keluvchi video va animatsiyalarni tanlaydi. Keyin esa Surdo imo-ishora tilining grammatik qoidalariga mos ravishda ifodalaydi. Asosiy texnologik yechim sifatida tabiiy tilni qayta ishlash moduli qo'llaniladi.

Kalit so'zlar: Imo-ishora tili, tabiiy tilni qayta ishlash, tokenizatsiya, lemmatizatsiya, sintaktik tahlil.

Har bir mamlakatda, shu jumladan O'zbekistonda ham, eshitish va nutq imkoniyati cheklangan shaxslar uchun asosiy muloqot vositasi bo'lgan o'ziga xos imo-ishora tili ya'ni Surdo imo-ishora tili mavjud. Afsuski, jamiyatning aksariyat qismi ushbu tilni bilmasligi yoki e'tiborsizligi sababli, oddiy insonlar va eshitish imkoniyati cheklangan insonlar o'rtasida chuqur kommunikatsion bo'shliq yuzaga kelgan. Bu bo'shliq nafaqat kundalik muloqotni qiyinlashtiradi, bu insonlarning ta'lim olish, sog'liqni saqlash va davlat xizmatlaridan foydalanish kabi muhim sohalarida teng imkoniyatlarga ega bo'lishiga to'sqinlik qiladi va ularni ijtimoiy hayotdan chetlashtirib quyadi.

Hozirgi kunda O'zbekistonda, Surdo imo-ishora tilidan foydalanish darajasi, ayniqsa rasmiy va ta'lim sohalarida, yetarli emas. Ta'lim tizimida va o'qitish jarayonlarda imo-ishora tiliga yetarlicha e'tibor qaratilmasligi hamda ota-onalarning Surdo imo-ishora tilining ahamiyatini to'liq anglab yetmasligi mavjud vaziyatni yanada murakkablashtirmoqda. Natijada, eshitish imkoniyati cheklangan shaxslarning jamiyatga to'liq integratsiyalashuvi va ularning salohiyatidan foydalanish imkoniyatlari sezilarli darajada cheklanmoqda.

Shu bilan birga, zamonaviy axborot texnologiyalari, xususan, tabiiy tilni qayta ishlash (NLP) va kompyuter ko'rishi (Computer Vision) sohalarining rivojlanishi eshitish va nutq imkoniyati cheklangan shaxslar uchun kommunikatsion to'siqlarni kamaytirish imkonini bermoqda. Imo-ishora tillarini avtomatik aniqlash va tarjima qilish tizimlari bu jarayonda muhim rol o'ynaydi. Eshitish va gapirish

imkoniyati cheklangan shaxslar bilan ikki tomonlama muloqot o'rnatish nafaqat aloqa imkoniyatlarini kengaytiradi, balki ularning jamiyat va texnologik taraqqiyotda faol ishtirok etishini ham ta'minlaydi [6].

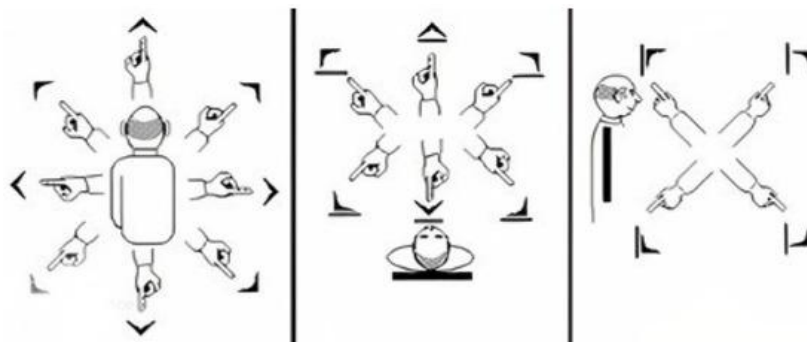
O'zbek surdo imo-ishora tili. Har bir mamlakatning imo-ishora tili o'z milliy tili va madaniyatiga asoslangan holda shakllanadi. O'zbek Surdo imo-ishora tili ham o'zbek tilining leksik va grammatik xususiyatlaridan ta'sirlangan bo'lsa-da, o'ziga xos strukturaga ega bo'lgan mustaqil til tizimidir. [1].

Imo-ishora tili eshitish imkoniyati cheklangan shaxslar uchun o'z fikrini erkin ifodalash, balki ijtimoiy hayotda faol ishtirok etish va to'laqonli hayot kechirish imkoniyatlarini yaratishda muhim rol o'ynaydi. Eshitish imkoniyati cheklanganlar jamiyatida Surdo imo-ishora tilidan keng foydalanilayotgan bo'lsa-da, rasmiy ta'lim tizimida uning qo'llanilishi hanuzgacha yetarli darajada rivojlanmagan. Bu masalalar oliy ta'lim muassasalarida kadrlarni tayyorlash dasturlarida Surdo imo-ishora tilidan samarali foydalanishga yetarlicha e'tibor berilmaganida aks etadi. Misol: O'qitish jarayonida eshitish va gapirish imkoniyati cheklangan o'quvchilarning ehtiyojlari to'liq inobatga olinmayapti. Eshitish imkoniyati cheklangan bolalarning ota-onalari ko'pincha imo-ishora tilining aloqa bo'shliqlarini bartaraf etishdagi ahamiyatini bilishmaydi[2].

Ushbu maqolada quyidagi masalalarga yechim topish maqsad qilingan:

- Eshitish va gapirish imkoniyati cheklangan shaxslarning muloqot imkoniyatlarini kengaytirish: Surdo imo-ishora tili orqali eshitish va gapirish imkoniyati cheklangan shaxslarning o'zaro va jamiyat bilan samarali muloqot qilishlariga ko'maklashish.
- Oliy va o'rta ta'lim muassasalarida eshitish imkoniyati cheklangan o'quvchilar va talabalar uchun Surdo imo-ishora tilidan foydalanib o'qitish shaklini qo'llash.
- Hukumat rahbarlari, o'qituvchilar, mutaxassislar, jamoat rahbarlari va jamoatchilik kabi turli guruhlarini Surdo imo-ishora tili so'zlashuvchilar bilan muloqotini yengillashtirish.

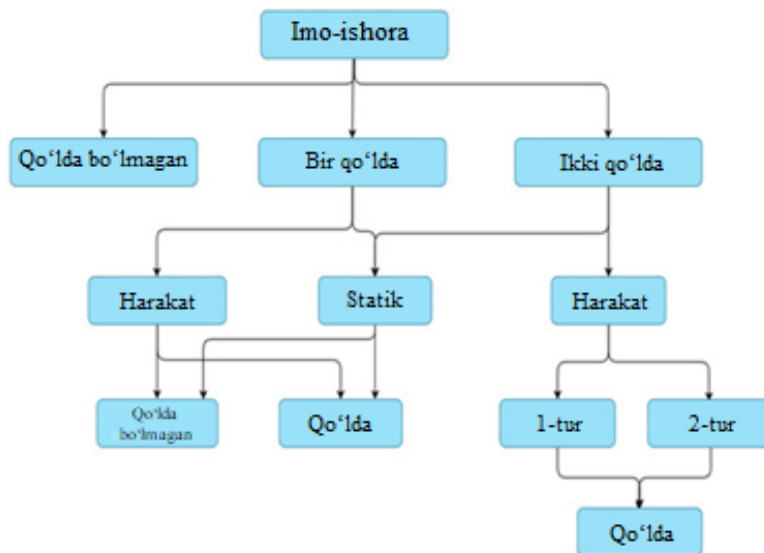
Surdo imo-ishora tili, 1-rasmda ko'rsatilganidek, imo-ishoralar va ko'rsatkichlarning turli shakllarini o'z ichiga oladi.



1-rasm. Imo-ishora tilidagi harakat va ifoda yo‘nalishlari.

Surdo imo-ishora tili faqat qo‘l imo-ishora harakatlari bilan cheklanmay, balki yuzning mimika ifodalari, tana holati va harakatlari kabi komponentlarni ham o‘z ichiga oladi. Ushbu komponentlar imo-ishoralarning ma’nosini kuchaytirish, gap ohangini ifodalash hamda grammatik bog‘liqlikni ifodalash kabi funksiyalarni bajarib beradi. Imo-ishora tilining chuqurroq tahlili asosida uni 2-rasmda tasvirlanganidek uchta asosiy toifaga ajratish mumkin.

Imo-ishora tili nafaqat qo‘l harakatlari, balki yuz ifodalari, tana holati, og‘iz va lablarning hamohang ifodalarini ham o‘z ichiga oladi. Ushbu elementlar imo-ishora tilining ajralmas qismlaridan biri bo‘lib, modulyatsiya vositasi sifatida xizmat qiladi.



2-rasm. Qo‘l imo-ishoralarining turlari.

Bizning og‘zaki nutqda ohang va intonatsiya orqali fikr ifodasi kuchaytirilib, aniqlashtirilsa, surdo imo-ishora tilida esa tana harakati va yuz ifodalari orqali ma’no kuchaytiriladi, gapning ohangi va kayfiyati aniqlik bilan yetkaziladi. Shu tariqa, tana imo-ishoralari muloqot jarayonida semantik yukni oshirish, emotsional fonni shakllantirish va grammatik tuzilmalarni aniqlashtirish vazifasini bajaradi[5].

Shuningdek bir qo'l bilan amalga oshiriladigan imo-ishoralar ikki toifaga ajratiladi:

- **Statik harakatlar** – belgining ma'lum bir holatda qotib turishi.
- **Dinamik harakatlar** – belgining harakatlanish jarayonida ifodalanishi.

Ushbu harakatlar qo'lga tegishli yoki qo'l bilan bog'liq bo'lmagan shakllarda amalga oshirilishi mumkin. Ikki qo'l ishtirokidagi imo-ishoralar ham statik va dinamik harakatlarga bo'linadi. Biroq, ular yana quyidagi turlarga tasniflanadi:

- **0-tur** – har ikkala qo'l teng darajada ishtirok etadigan imo-ishoralar.
- **1-tur** – ikkala qo'l ishtirok etadigan, biroq ulardan biri yetakchi, ikkinchisi yordamchi vazifani bajaradigan imo-ishoralar.

Mazkur tasnif imo-ishora tilidagi murakkablik va ifoda boyligini chuqurroq anglash imkonini beradi hamda uni o'rganish va o'qitish jarayonida muhim metodologik asos vazifasini bajaradi.

O'zbek Surdo imo-ishora tili grammatikasi o'zbek tilining grammatik tizimidan sezilarli darajada farqlanishi mumkin [1]. Xususan:

- **So'z tartibi:** Surdo imo-ishora tilida gap tuzilishi ko'pincha *Ega-Olmosh-Fe'l* strukturasi asosida tashkil topadi.
- **Fe'llarning tuslanishi:** Fe'l shakllari imo-ishoralarda kontekstual ravishda ifodalanadi va ularning zamon, shaxs yoki nisbat shakllari alohida grammatik belgilar orqali emas, balki imo-harakat va yuz ifodalari vositasida bildiriladi.
- **Ko'plik va egalik shakllari:** O'zbek tilidagi kabi ko'plik yoki egalik qo'shimchalari imo-ishoralarda aniq grammatik affikslar bilan emas, balki maxsus harakatlar, ko'rsatkichlar yoki qo'shimcha mimika orqali ifodalanadi.

Shu sababli, og'zaki nutqdan imo-ishora tiliga so'zma-so'z tarjima qilish ko'pincha noto'g'ri yoki tushunarsiz natijalarga olib keladi. Tarjima jarayonida O'zSITning o'ziga xos grammatik xususiyatlarini e'tiborga olish zarur. Aks holda, muloqotda semantik buzilishlar va tushunmovchiliklar yuzaga kelishi mumkin.

Taklif etilayotgan usul: Ushbu maqolada taqdim etilgan taklif qilingan tizim eshitish qobiliyati zaif odamlarga boshqa eshitish qobiliyatiga ega odamlar bilan osongina muloqot qilishda yordam beradigan real vaqt rejimida nutqni Surdo imo-ishora tiliga o'tkazish tizimidir. Tizim asosan oltita komponentdan iborat:

1. Kirishni qayta ishlash moduli – foydalanuvchi tomonidan kiritilgan ma'lumotni tizim uchun mos shaklga keltirish vazifasini bajaradi. Ushbu modul ikki asosiy funksiyani o'z ichiga oladi:

- **Nutqni matnga aylantirish (Speech-to-Text, STT)** - Agar foydalanuvchi ma'lumotni ovozli nutq shaklida kiritsa, birinchi bosqichda nutqni matnga

aylantirish jarayoni amalga oshiriladi. Bunda o'zbek tiliga moslashtirilgan maxsus Speech-to-Text (STT) modelidan foydalanish talab etiladi. STT tizimining aniqligi butun tarjima tizimi samaradorligiga bevosita ta'sir ko'rsatadi. Biroq, muloqot jarayonidagi fon shovqinlari, turli hudud dialektlari va individual talaffuz xususiyatlari STT modulining ishlash sifatiga sezilarli darajada ta'sir qilishi mumkin [2].

- *To'g'ridan-to'g'ri matn kiritish* - Foydalanuvchi ma'lumotni bevosita matn shaklida kiritishi ham mumkin. Bunday holatda, Speech-to-Text bosqichi o'tkazib yuboriladi va tizim matn bilan bevosita ishlay boshlaydi.

2. Tabiiy tilni qayta ishlash (NLP) moduli - tizimning yadrosi bo'lib, foydalanuvchi tomonidan kiritilgan matnni lingvistik jihatdan tahlil qiladi va uni Surdo imo-ishora tiliga tarjima qilish uchun tayyorlaydi (4-rasmga qarang).

Ushbu modul quyidagi asosiy bosqichlardan iborat [2]:

- *Dastlabki ishlov berish* — Matn keraksiz belgilardan (masalan, tinish belgilari va ortiqcha bo'shliqlardan) tozalanadi va normallashtiriladi. Masalan, barcha harflar kichik harf shakliga keltiriladi, sonlar esa so'z shaklida ifodalanadi. Ushbu bosqich matnni bir xil va tizimli ko'rinishga keltirishga xizmat qiladi.
- *Tokenizatsiya* — Matn alohida birliklarga — tokenlarga (odatda so'zlar yoki gap bo'laklariga) ajratiladi. O'zbek tilida so'z o'zagiga ko'plab qo'shimchalar qo'shilishi mumkinligi sababli, dastlab so'z darajasidagi tokenizatsiya amalga oshiriladi.
- *Nomuhim so'zlarni olib tashlash (Stopword Removal)* — Matnda tez-tez uchraydigan, lekin semantik yuklamasi past bo'lgan so'zlar (masalan, "va", "bilan", "uchun", "bu", "shu") aniqlanib, olib tashlanadi. Bu jarayon tarjima sifatini oshirish va asosiy ma'noli birliklarga e'tiborni jamlash uchun muhimdir.
- *Stemming va Lemmatizatsiya* — Har bir so'zning morfologik tahlili amalga oshiriladi va u lug'aviy asosiy shakliga (lemma) keltiriladi. Masalan, "kitoblarimni" so'zi "kitob" lemmaga qisqartiriladi. Ushbu bosqich imo-ishoralar ma'lumotlar bazasi bilan mos ishlashni ta'minlash uchun zarur.
- *Nutq qismlarini belgilash (POS-Tagging)* — Har bir tokenning nutqdagi turkumi (masalan, ot, fe'l, sifat, ravish va boshqalar) aniqlanadi. Ushbu ma'lumot keyingi bosqichda sintaktik tahlil va grammatik qoidalarni to'g'ri qo'llash uchun asos bo'lib xizmat qiladi.
- *Sintaktik tahlil (Parsing)* — Gapning grammatik strukturasini aniqlash jarayoni bo'lib, bunda so'zlar o'rtasidagi munosabatlar va frazaviy birliklar

tahlil qilinadi. Bu bosqich imo-ishora tiliga mos grammatik tuzilma hosil qilish uchun muhim ahamiyat kasb etadi.

3. Tarjima va natijani generatsiya qilish moduli - Taklif etilayotgan tizim Surdo imo-ishora tiliga nutq va matn asosida avtomatik tarjima qilish uchun mo'ljallangan bo'lib, quyidagi asosiy bosqichlardan iborat [3]:

- *Qoidalariga asoslangan tarjima qilish* orqali kiritilgan matn imo-ishora tili grammatik tuzilmasiga mos ravishda so'zlar ketma-ketligi qayta tartiblanadi.
- *So'zlarni aniqlash* mexanizmi yordamida barqaror frazalar aniqlanib, agar mavjud bo'lsa, yagona imo-ishora belgisi bilan ifodalanadi.
- *Ma'lumotlar bazasidan qidirish* orqali so'zlar yoki iboralarga mos imo-ishora videolari yoki animatsiyalari topiladi.
- *Zaxira mexanizmi* sifatida mos keluvchi imo-ishora topilmaganda, barmoq alifbosi yordamida harflab ko'rsatish funksiyasi qo'llaniladi.
- *Natijani generatsiya qilish* orqali tanlangan imo-ishora videolari va animatsiyalar imo-ishora tili grammatik qoidalariga mos tartibda birlashtirilib, foydalanuvchiga taqdim etiladi. Animatsiyalar 3D avatar texnologiyalari yordamida yuz ifodasi va tana harakati kabi imo-ishoralardan orqali ham ifodalanadi.

Xulosa. Ushbu maqolada o'zbek tilidagi nutq va matnni Surdo imo-ishora tiliga tarjima qiluvchi tizim konsepsiyasi va metodologiyasi taklif etildi. Tizim NLP bosqichlari tokenizatsiya, lemmatizatsiya, sintaktik hamda imo-ishora tili grammatik qoidalarini qo'llash orqali foydalanuvchi kiritgan ma'lumotni mos ravishda imo-ishora ketma-ketligiga aylantiradi. Natijada, foydalanuvchi tomonidan kiritilgan matnning mazmuni ma'lumotlar bazasidan tanlangan video yoki animatsiyalar yordamida ifodalanadi.

Taklif etilayotgan tizim eshitish va nutq qobiliyati cheklangan shaxslar uchun ta'lim, kundalik hayot va davlat xizmatlarida muloqot imkoniyatlarini sezilarli darajada kengaytirishga xizmat qiladi. Shu bilan birga, sifatli imo-ishora ma'lumotlar bazasini yaratish, o'zbek tili uchun NLP vositalarini takomillashtirish, imo-ishora tili grammatikasini standartlashtirish va imo-ishora tili elementlarni tabiiy tarzda hosil qilish tizim rivojlanishining asosiy muammolari sifatida ko'riladi.

Kelgusida tizimning dastlabki prototipini ishlab chiqish, uni real foydalanuvchilarda sinovdan o'tkazish va natijalarga asoslangan takomillashtirish ishlari amalga oshirilishi rejalashtirilgan. Tadqiqot natijalari O'zbekistonda inklyuziv jamiyatni rivojlantirish va kommunikatsion to'siqlarni texnologiyalar orqali kamaytirishga hissa qo'shishi kutilmoqda.

Foydalanilgan adabiyotlar:

1. Jurayev D.B, Ochilov M.M, Abdullayeva M.I. Nutqni imo-ishora tiliga tarjima qilish tizimlarida matnlarga ishlov berish. TATU xabarlar, 4/68(2023).
2. Mukhammadjon Musaev, Dilshod Juraev, Mannon Ochilov, Malika Abdullaeva. “Text processing in speech translation systems into sign language”. International Scientific Conference on Modern Problems of Applied Science and Engineering MPASE2024, Samarkand, Uzbekistan • 2–3 May 2024, AIP Conf. Proc. 3244, 030062 (2024) <https://doi.org/10.1063/5.0241738>
3. Zhu, D., Czehmann, V., & Avramidis, E. (2023). Neural Machine Translation Methods for Translating Text to Sign Language Glosses. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 12523–12541. <https://aclanthology.org/2023.acl-long.700/>
4. Moryossef, A., Jiang, Z., Müller, M., Ebling, S., & Goldberg, Y. (2023). Linguistically Motivated Sign Language Segmentation. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 12703–12724. <https://doi.org/10.18653/v1/2023.findings-emnlp.846>
5. Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation. <https://arxiv.org/abs/2003.13830>
6. Shahin, N., & Ismail, L. (2024). From Rule-Based Models to Deep Learning Transformers Architectures for Natural Language Processing and Sign Language Translation Systems: Survey, Taxonomy and Performance Evaluation. <https://arxiv.org/abs/2408.14825>