# O‘ZBEK TILIDA GRAFEMA-FONEMA (G2P) MODELI YARATISHNING AMALIY ASOSLARI

**Maxmudjonova Gulshaxnoz Ulug‘bek qizi**
*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti*
*Kompyuter lingvistikasi va raqamli texnologiyalar kafedrasi tayanch doktoranti*

**DOI:** https://doi.org/10.5281/zenodo.15179399

***Annotatsiya****: Zamonaviy tilshunoslik va kompyuter lingvistikasi rivojlanishi bilan matndan nutqqa (Text-to-Speech, TTS) tizimlarini yaratishda grafema-fonema (G2P) konversiyasi muhim ahamiyat kasb etmoqda. O‘zbek tili kabi kam resursli tillar uchun yuqori sifatli TTS tizimlarini yaratish uchun G2P modellarini ishlab chiqish zarur. Ushbu maqolada O‘zbek tili uchun G2P modelini yaratishning nazariy asoslari va amaliy yondashuvlari tahlil qilinadi.*

***Kalit so‘zlar****: fonetik transkripsiya, fonema, grafema, fonologiya, nutq sintezatori.*

G2P modeli grafema (yozma belgilar) va fonema (talaffuz birliklari) o‘rtasidagi moslikni aniqlashga asoslanadi. O‘zbek tilining morfofonologik xususiyatlari, urg‘u tizimi, fonetik o‘zgarishlar modelining asosiy elementlari hisoblanadi. G2P modeli ishlab chiqish tilshunoslikning amaliy masalalaridan biri bo‘lib, yozma matnni og‘zaki nutqqa aylantirish jarayonida ishlatiladi. Bu quyidagi muammolarni hal qilishga yordam beradi: *Matndan nutqqa tizimlari[1]* – avtomatik ovozli assistentlar, ovozli tarjimonlar va nutq sintezatorlari uchun muhim texnologiya hisoblanadi. Ushbu tizimlar foydalanuvchilarga matnni real vaqt rejimida ovozli formatga o‘tkazish imkonini beradi. Bundan tashqari, G2P modeli raqamli yordamchilar, ta’lim platformalari, ko‘rish qobiliyati cheklangan insonlar uchun maxsus dasturlar va audiokitoblar yaratishda keng qo‘llaniladi. *Nutqni aniqlash va transkripsiyalash* – avtomatik subtitr yaratish, nutqni tahlil qilish kabi sohalarda qo‘llaniladi. *Kam resursli tillarni raqamlashtirish* – O‘zbek tili kabi tillarning fonetik qoidalarini aniq aniqlash va to‘g‘ri talaffuz qilish imkoniyatini beradi. Kam resursli tillar uchun Grafema-Fonema va Fonema-Grafema konversiyalarini yaxshilash uchun sintetik o‘quv ma’lumotlaridan foydalanishni yaxshi natija beradi. Kam resursli tillarda mavjud fonetik transkripsiyalar yetarli bo‘lmagani uchun, tadqiqotchilar [2] sun’iy yo‘l bilan yaratilgan ma’lumotlar yordamida modellarni yaxshilashni taklif qiladilar. *O‘qish va yozishni o‘rganish jarayonlari* – til o‘rganish tizimlari uchun yordamchi bo‘lib xizmat qiladi.

O‘zbek tili uchun G2P modelini yaratishda asosiy muammolar: ***Kam resurslilik*** – O‘zbek tilida fonetik transkripsiyalar va lug‘atlar yetarli emas. ***Orfoepik xususiyatlar*** – O‘zbek tilida talaffuz ba’zan yozuvdan farq qiladi (masalan, “bo‘l” so‘zi “bŭl” tarzida talaffuz qilinishi). ***Leksik o‘zgaruvchanlik*** – Qat’iy fonetik

qoidalarga ega bo‘lmagan ayrim so‘zlar mavjud (arabcha yoki forscha so‘zlar talaffuzi turlicha bo‘lishi mumkin).

*O‘zbek tili uchun G2P modelini yaratishning amaliy bosqichlari:*

**Fonemik tahlil** – fonema to‘plamini aniqlash va urg‘u qoidalarini belgilash. G2P modelining asosiy vazifasi – harf (grafema) va fonema o‘rtasidagi bog‘liqlikni aniqlash**.** Buni to‘g‘ri bajarish uchun avvalo fonemik tahlil qilish kerak. Shuningdek, o‘zbek tilidagi fonemalar to‘plamini aniqlash.

**Ma’lumotlar yig‘ish va tayyorlash** – O‘zbek tilidagi matn va fonetik transkripsiyalar bazasini yaratish. Fonetik transkripsiya bazasini yaratish uchun, birinchi navbatda o‘zbek tilida fonetik lug‘at yaratish kerak. G2P modelining [3] sifati katta hajmdagi fonetik lug‘at va transkripsiya ma’lumotlariga bog‘liq. Shuning uchun, O‘zbek tili uchun fonetik lug‘at yaratish quyidagi bosqichlarda amalga oshiriladi. Manbalarni yig‘ish. O‘zbek tilidagi so‘zlar ro‘yxatini tuzish (100,000+ so‘z) va mavjud orfoepik lug‘atlardan foydalanish[4]. Shuningdek, turkumlangan (morfologik) lug‘atlardan foydalanish. Fonetik qoidalarni aniqlash. O‘zbek tilidagi asosiy unli va undosh tovushlar ro‘yxatini yaratish. Orfoepik o‘zgarishlarni aniqlash (masalan, “ketdi” → “ketti”). Assimilyatsiya va regressiv o‘zgarishlarni aniqlash.

**Fonetik transkripsiya yaratish** – Har bir so‘z uchun qoida asosida fonetik transkripsiya yaratish. Ushbu transkripsiyalarni fonetik belgilar bilan yozish (IPA yoki X-SAMPA formatida) va avtomatlashtirilgan fonetik transkripsiya tizimi yaratish. Dastlabki transkripsiya qoidalarini yozish (eSpeak NG yordamida). Modellashtirish uchun mashina o‘rganish usullaridan foydalanish.

**Sinov va baholash** – modelning aniqligini fonetik transkripsiyalar bilan solishtirib baholash.

**Natijalarni optimallashtirish** – modelni yaxshilash uchun qoidalar yoki neyron tarmoqlar tuzilmasini o‘zgartirish.

Morfologik lug‘atlardan foydalanish[2] va ularning G2P modelida ahamiyati. Morfologik lug‘atlar — bu lug‘atlar har bir so‘z shaklini nafaqat asosiy (leksik) shaklda, balki uning grammatik xususiyatlari (turkumi, zamon, son, egalik, kelishik, fe’l shakllari) bilan birga taqdim etadigan ma’lumotlar bazasidir. G2P modelida morfologik lug‘atlardan foydalanish modelga so‘zlarning morfologik xususiyatlarini tushunishga va aniq transkripsiya yaratishga yordam beradi.

**Natijalar va xulosa.** Tadqiqot natijalariga ko‘ra, O‘zbek tili uchun yaratilgan G2P modeli fonetik aniqlikni oshirishga yordam beradi va TTS tizimlari sifatini yaxshilashga xizmat qiladi. G2P modelining tilshunoslikning amaliy masalalariga kiritilishi uning nutq texnologiyalaridagi muhim roli bilan bog‘liq. Kelajakda fonetik

bazani kengaytirish va neyron tarmoqlardan yanada samarali foydalanish orqali modelning imkoniyatlarini yanada kengaytirish rejalashtirilmoqda.

**Foydalanilgan adabiyotlar:**

1. Paul Taylor. Hidden Markov Models for grapheme to phoneme conversion. In Proceedings of the 9th European Conference on Speech Communication and Technology, 2005
2. S. Lauly, R. Collobert, A. Bouchekam, and P. Krasheninnikov, "Low-Resource G2P and P2G Conversion with Synthetic Training Data," in *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 2020, pp. 108–117. Available: https://aclanthology.org/2020.sigmorphon-1.12.pdf.
3. Nurmonov.A Sobirov.A Qosimova.N Hozirgi o,,zbek adabiy tili. – Toshkent, 2013
4. Rahmatullayev Sh. O'zbek tili morfem lug'ati. – Toshkent, 1977.
5. Mamatov.J Chiniqulov O'zbek tilining orfoepik lug'ati Toshkent, 2021