

AUDIOMATNLARDA MORFOLOGIK TAHLIL BOSQICHI

Urazaliyeva Mavluda Yangiboyevna,
mustaqil izlanuvchisi
urazaliyevam@nuu.uz
O'zbekiston Milliy universiteti

Annotatsiya. Ushbu maqolada audiomatnlar asosida o'zbek tilining morfologik tahlil bosqichlarini chuqur o'rganishga qaratilgan lingvistik va texnologik yondashuvlar tahlil qilinadi. Tadqiqotda audiomatnlarning murakkab fonetik va sintaktik xususiyatlari hisobga olingan holda, avtomatik morfemik segmentatsiya, leksik analiz, grammatik kategoriya aniqlash va kontekstual morfologik disambiguatsiya bosqichlari alohida ilmiy metodologiya asosida bayon etilgan. Xususan, BERT arxitekturasiga asoslangan chuqur o'rgatish modellarining o'zbek tilidagi og'zaki nutq korpusi (uzbekcorpus.uz hamda kun.uz transkriptlari) asosida sinovdan o'tkazilishi natijasida yuqori aniqlikka ega morfologik tahlil tizimining yaratilishi nazariy va amaliy jihatdan asoslab berilgan.

Abstract. This article analyzes linguistic and technological approaches aimed at an in-depth study of the morphological analysis stages of the Uzbek language based on audio texts. The research describes the stages of automatic morphemic segmentation, lexical analysis, grammatical category determination, and contextual morphological disambiguation, taking into account the complex phonetic and syntactic features of audio texts, based on a distinct scientific methodology. In particular, the development of a high-precision morphological analysis system has been theoretically and practically substantiated through testing deep learning models based on the BERT architecture using the Uzbek language oral speech corpus (transcripts from uzbekcorpus.uz and kun.uz).

Аннотация. В данной статье анализируются лингвистические и технологические подходы, направленные на углубленное изучение этапов морфологического анализа узбекского языка на основе аудиотекстов. В исследовании описываются этапы автоматической морфемной сегментации, лексического анализа, определения грамматических категорий и контекстуальной морфологической дисамбигуации с учетом сложных фонетических и синтаксических особенностей аудио текстов на основе специальной научной методологии. В частности, разработка высокоточной системы морфологического анализа была теоретически и практически обоснована путем тестирования моделей глубокого обучения на базе архитектуры BERT с использованием корпуса устной речи узбекского языка (транскрипции с uzbekcorpus.uz и kun.uz).

Kalit soʻzlar: *audiomatn, morfologik tahlil, fonetika, segmentatsiya, BERT, tabiiy tilni qayta ishlash (NLP), stilistika.*

Zamonaviy tilshunoslik, ayniqsa raqamli lingvistika va avtomatik nutqni qayta ishlash yoʻnalishlarida, audiomatnlarni tahlil qilish alohida ahamiyatga ega. Bunda nutq birligi sifatida shakllangan ogʻzaki matnlar lingvistik struktura, grammatik qurilish va semantik qatlam nuqtai nazaridan chuqur tahlil qilinadi. Ayniqsa, avtomatik nutqni aniqlash (ASR – Automatic Speech Recognition) va tabiiy tilni qayta ishlash (NLP – Natural Language Processing) tizimlari uchun audiomatnlarning grammatik strukturasini aniq belgilash texnologik jarayonning poydevoridir.

Audiomatnlarni morfologik jihatdan tahlil qilish ularning grammatik tuzilmasini aniqlashda muhim bosqich hisoblanadi. Bu bosqichda audiomatnlar transkripsiya qilinib, soʻzlar morfemalarga ajratiladi va har bir morfema grammatik kategoriyalar nuqtai nazaridan belgilab chiqiladi.

Shu oʻrinda qayd etish lozimki, oʻzbek tili agglutinativ til boʻlgani sababli morfologik tahlilda affiksial koʻrsatkichlarning roli beqiyosdir. Audiomatnlarda morfologik tahlil bosqichi ularning tarkibiy qismlarini aniqlash, grammatik birliklar va ularning oʻzaro munosabatlarini koʻrsatish, shu bilan birga morfemik strukturalarni avtomatlashtirilgan tarzda ajratishga xizmat qiladi. Bu jarayon oʻzbek tilining lingvistik xususiyatlarini chuqur aks ettirib, til korpuslarini boyitishga hamda sunʼiy intellekt asosidagi til modellari uchun mustahkam asos yaratadi [1].

Morfologik tahlil natijalari asosida korpusga morfologik teglar biriktiriladi. Bu esa til korpusining yanada boyitilishiga va avtomatik analiz tizimlarining takomillashtirilishiga xizmat qiladi.

Audiomatnlarda morfologik tahlil bosqichi — tilshunoslik, kompyuter lingvistikasi va sunʼiy intellekt asosidagi tizimlar uchun til materialining grammatik va strukturaviy xususiyatlarini aniqlovchi eng muhim bosqichlardan biridir. Ushbu tahlilning asosiy vazifasi — ogʻzaki nutqda shakllangan til birliklarini transkripsiya qilish, ularni morfemalarga ajratish va grammatik kategoriyalar asosida teglash orqali ularning funksiyalarini aniqlashdir.

Birinchi bosqich *transkripsiya* boʻlib, audiomatn fonetik shakldan ortografik (yozma) shaklga koʻchiriladi. Bu jarayonda real talaffuzdagi ohanglar, urgʻular, qisqarishlar, eliziyalar va boshqa artikulyator xususiyatlar inobatga olinadi. Masalan, uzbekcorpus.uz platformasidagi audiomatnlarda “ketayotgan” feʼlining ogʻzaki shakli [ketʼyotgan] tarzida transkripsiya qilinadi. Nutq tezligi, urgʻu oʻzgarishi va regional talaffuz bu bosqichga bevosita taʼsir qiladi [1]. Shu bois, transkripsiyalashda avtomatik vositalardan tashqari lingvistik ekspertizaga asoslangan tekshiruv ham zarur.

Bugungi kunda transkripsiyani avtomatlashtirish bo'yicha jahon tajribasida bir nechta ilg'or modellar mavjud. Masalan, Mozilla'ning DeepSpeech, Metaning wav2vec 2.0 yoki Kaldi kabi ochiq kodli ASR (Automatic Speech Recognition) modellar audiomatnni yozma matnga aylantirishda keng qo'llaniladi [2]. Ushbu modellar o'zbek tiliga moslashtirilsa, transkripsiya jarayoni yanada avtomatlashtirilgan va sifatli amalga oshirilishi mumkin.

Transkripsiyadan keyingi bosqich *morfemik tahlil* hisoblanadi. Bu bosqichda matndagi har bir so'z grammatik birliklarga — morfemalarga — ajratiladi va ular grammatik kategoriya bo'yicha belgilanadi. Masalan, “qilolmayapti” fe'li uzbekcorpus.uz korpusida qil+OL+NEG+PROG+3SG tarzida teglanadi: OL — imkon, NEG – inkor, PROG – hozirgi davomiy zamon, 3SG – uchinchi shaxs, birlik[3]. Bu kabi tahlillar [4] affiksial tizim tasnifiga to'liq mos keladi.

Yana bir muhim jihat shundaki, fonetik jihatdan qisqargan shakllar morfemik tahlil jarayonida noto'g'ri ajratilish xususiyatiga ega. Masalan, “kelayotgan edi” so'z birikmasi og'zaki shaklda [kelyatkanidi] tarzida transkripsiya qilinadi. Bunday hollarni aniqlash uchun fonetik-morfologik segmentatsiya modellari (Foneme2Morph) ishlab chiqilmoqda. Ular tovushlar ketma-ketligidan morfemalarni ajratib olish imkonini beradi[5].

Audiomatnlarda morfologik tahlil ikki komponent — fonetik va grammatik komponentlarning uyg'un ishlashini talab qiladi. Bu uyg'unlik transkripsiyadan boshlab teglashgacha bo'lgan barcha bosqichlarda aniqlik, tizimlilik va tilshunoslikka asoslangan yondashuvni talab etadi.

O'zbek tili turkiy tillar oilasiga mansub bo'lib, agglutinatив tillar qatoriga kiradi. Agglutinatивlik hodisasi orqali so'z asoslariga grammatik ma'no ifodalovchi affikslar qatma-qat qo'shib, yangi shakllar va ma'nolar hosil bo'ladi hamda bunda so'z negizlari o'zining shakliy va ma'noviy mustaqilligini saqlab qoladi. Har bir affiks o'zining alohida grammatik funksiyasiga ega bo'lib, mustaqil morfema sifatida ko'riladi[4].

Agglutinatив tillarning asosiy xususiyati shundan iboratki, har bir qo'shimcha faqat bitta grammatik ma'noni ifodalaydi va ularni so'z tarkibida aniq farqlash mumkin; bu esa morfologik tahlil jarayonini ancha soddalashtiradi. Ushbu jihatdan agglutinatив tillar flektiv tillardan farqlanadi, chunki flektiv tillarda bir qo'shimcha bir nechta grammatik ma'noni ifodalash xususiyatiga ega bo'ladi va morfemalar orasidagi chegaralar unchalik aniq emas[6].

Masalan, “boryapman” fe'l shaklini morfologik tahlil qiladigan bo'lsak, quyidagi morfemalarni ajratish mumkin: bor — fe'l asosi (leksik morfema); -ya — davomiylik morfemasi; -p — shaxs qo'shimchasidan oldingi fonetik ko'makchi; -man — 1-shaxs, birlik ko'rsatkich morfemasi.

O'zbek tilining agglutinativlik xususiyatini kuzatishda uzbekcorpus.uz platformasida mavjud misollardan foydalanish mumkin bo'lib, bu til materiallari va ularning tahlillarini yanada boyitishga yordam beradi:

“O'qiyotganimda” so'z shakli quyidagicha tahlil qilinadi: o'qi (fe'l asosi), -yot (davomiylilik), -gan (sifatdosh yasovchi qo'shimcha), -im (egalik), -da (kelishik qo'shimchasi).

“Bilimdonligimizdan” shakli: bilim (ot), -don (sifat yasovchi qo'shimcha), -lik (abstrakt ot yasovchi qo'shimcha), -imiz (egalik ko'plik shakli), -dan (kelishik).

“Mustaqilligimizning” shakli: mustaqil (sifat), -lik (abstrakt ot yasovchi), -imiz (egalik), -ning (qaratqich kelishik).

Ushbu xususiyat tufayli o'zbek tilida so'z hosil qilish imkoniyatlari kengayib, lug'at tarkibi boyitiladi hamda grammatik shakllar aniq ifodalangan holda yuzaga keladi, shuningdek, so'z yasash jarayonida aniqlik va tizimlilik ta'minlanadi. Jahon tilshunosligida ham agglutinativ tillarning morfologik tuzilishi va xususiyatlari bo'yicha ko'plab tadqiqotlar olib borilgan bo'lib, bu tadqiqotlar morfemalar orasidagi munosabatlar va ularning morfologik strukturalarini batafsil ochib beradi [7].

Demak, o'zbek tilining agglutinativ tabiati uning morfologik tizimini boyitib, tilning ekspressivligini va so'z hosil qilishdagi imkoniyatlarini oshiruvchi muhim omil sifatida xizmat qiladi.

O'zbek tilida affiksalar morfemalar so'z shakllariga grammatik ma'no yuklashda muhim vosita hisoblanadi. Ular tilshunoslikning morfologiya bo'limida markaziy o'rin tutadi. Affiksalar ko'rsatkichlar yordamida so'zlarning zamon, shaxs, son, egalik, kelishik kabi grammatik kategoriyalarga mansubligi belgilanadi [8].

Birinchidan, affikslar orqali **zamon** kategoriyasi ifodalanadi. Masalan, fe'lga qo'shilgan -di, -moqda, -adi, -yapti kabi affikslar harakatning vaqt shaklini bildiradi: *bordim, ketmoqda, boradi, kelayapti* kabi.

Ikkinchidan, affiksalar morfemalar **shaxs va sonni** ifodalashda xizmat qiladi. Masalan: *keladi* (3-shaxs, birlik), *kelamiz* (1-shaxs, ko'plik), *kelasanlar* (2-shaxs, ko'plik) shakllarida bu grammatik ma'no affikslar orqali beriladi [9].

Uchinchidan, **egalik kategoriyasi** o'zbek tilida egaga tegishlilikni bildiruvchi affikslar yordamida ifodalanadi: *kitobim* (1-shaxs, birlik), *kitobing* (2-shaxs, birlik), *kitoblari* (3-shaxs, ko'plik) kabi shakllar buning yaqqol namunasi [10]. Shuningdek, **kelishik kategoriyasi** ham affikslar orqali amalga oshiriladi: *kitobni* (tushum), *kitobga* (jo'nalish), *kitobdan* (chiqish) va hokazo.

Affikslarning o'zaro kombinatsiyalanishi esa grammatik jihatdan murakkab sintaktik tuzilmalarni yaratishga imkon beradi. Masalan, *yozayotgan edim* shaklida

-ayotgan davomiylikni, -edi o'tgan zamonni, -m esa 1-shaxs birlikni bildiradi [11]. Bu kabi murakkab morfologik birliklar, nutqiy faoliyatda ifoda aniqligi va ma'no aniqligini ta'minlaydi.

O'zbek tilidagi affiksalar ko'rsatkichlar zamon, shaxs, son, egalik va kelishik kategoriyalarini shakllantirish va ifodalashda asosiy grammatik vosita hisoblanadi. Bunday tahlillar uzbekcorpus.uz platformasida mavjud audiomatnlarni misolida yanada chuqurroq tadqiq qilinishi mumkin bo'lib, amaliy tilshunoslikda muhim ahamiyatga ega.

Til korpuslari hozirgi zamon lingvistik tadqiqotlarining muhim qismi sifatida tilshunoslik sohasida keng qo'llanib kelmoqda. Har bir morfemalangan so'z teglar bilan belgilanishi morfologik tahlilning avtomatlashtirilishi uchun asos hisoblanadi. Ushbu jarayon avtomatik tizimlar orqali amalga oshirilishi esa korpuslarning samarali boyitilishi va leksik-semantik tahlillar sifatining oshishiga xizmat qiladi [12].

Morfologik teglash tizimi so'zlarning tarkibiy qismlariga ajratilib, morfemalar darajasida tahlil qilinishiga asoslangan bo'lib, har bir ajralgan morfema maxsus morfologik teglar bilan belgilab chiqiladi. O'zbek tilida ishlab chiqilgan morfologik teglar tizimi jahon miqyosida keng tarqalgan Universal Dependencies (UD) [13] formatiga asoslangan holda moslashtirilmoqda. Bu esa xalqaro lingvistik standartlarga mos keluvchi, shuningdek, o'zbek tilining morfologik xususiyatlarini to'liq aks ettiruvchi tizim yaratilishini ta'minlaydi.

Masalan, quyidagi shaklni ko'rib chiqaylik:

borayapman → bor+VA+PROG+1SG

Bu yerda:

VA — fe'l asos;

PROG — hozirgi davomiy zamon shakli;

1SG — birinchi shaxs, birlik shakli.

Bu kabi teglash amaliyotining afzalligi shundaki, murakkab grammatik tuzilishlar avtomatik tarzda aniqlanib, tizim yordamida morfologik jihatdan to'g'ri segmentatsiyalanadi. Bu o'z navbatida, so'zlarning grammatik va leksik xususiyatlarini chuqurroq tahlil qilish imkoniyatini yaratadi.

O'zbek tilida morfologik tahlil uchun quyidagi boshqa misollarni ham keltirish mumkin:

“yozmoqda” → yoz+VA+PROG+3SG

“kelishmoqda” → kel+VA+PROG+3PL

“o'qidim” → o'qi+VA+PST+1SG

uzbekcorpus.uz platformasi asosida olingan quyidagi misollar ham buni tasdiqlaydi:

“Gapirib turibman” → gapir+VA+CONV+tur+VA+PROG+1SG

“Ko‘rib chiqamiz” → ko‘r+VA+CONV+chiq+VA+FUT+1PL

“O‘ylab ko‘raman” → o‘y+VA+CONV+ko‘r+VA+PRES+1SG

Morfologik tahlilning avtomatik tarzda amalga oshirilishi tabiiy tilni qayta ishlash (NLP) sohasidagi ko‘plab vazifalarni (masalan, sintaktik va semantik tahlillar, matnlarni avtomatik ravishda klassifikatsiya qilish va boshqalar) bajarishni ancha osonlashtiradi. Bu jihatdan qaralganda, avtomatik teglash tizimlarining til korpuslari yaratish va rivojlantirishdagi roli nihoyatda katta [14].

Bundan tashqari, morfologik teglash yordamida turli murakkab tuzilmalar va shakllar avtomatik tarzda aniqlanadi, bu esa tilni o‘rganish, mashinali tarjima tizimlarini rivojlantirish va nutqni tanish texnologiyalarini yaratishda muhim imkoniyatlar ochadi [15]. Jumladan, o‘zbek tilida qo‘llanilayotgan zamonaviy morfologik tahlil tizimlari lemmatizatsiya (asos shaklini aniqlash) va annotatsiya jarayonlarini samarali tarzda bajarishga imkon beradi.

Shunday qilib, morfologik teglar va avtomatik teglash tizimlaridan foydalanish o‘zbek tilining lingvistik tadqiqotlarida xalqaro standartlarga muvofiqlikni ta‘minlash bilan birga, tahlil sifatini oshirish va tilni chuqurroq tadqiq etishga xizmat qiladi.

Zamonaviy lingvistika fanida til tizimining turli darajalari o‘zaro uzviy bog‘liqlikda o‘rganiladi. Ayniqsa, fonetika va morfologiya o‘rtasidagi o‘zaro aloqadorlik til birliklarini aniqlash, tahlil qilish va sun‘iy intellekt asosidagi til texnologiyalarini ishlab chiqishda muhim ahamiyat kasb etadi. Og‘zaki nutqda yuz beradigan fonetik o‘zgarishlar, aksariyat hollarda, morfemalarning shaklan buzilishiga yoki noto‘g‘ri talqin qilinishiga sabab bo‘ladi. Bu holat ayniqsa morfemik segmentatsiya, avtomatik transkripsiya va lemmatizatsiya jarayonlarida ko‘zga tashlanadi. Fonetik variantliklar, ya‘ni so‘z yoki so‘z birikmalarining talaffuzda fonologik shakli buzilishi, morfemik chegaralarni aniqlashni murakkablashtiradi.

Tilning fonetik va morfologik qatlamlari o‘zaro uzviy bog‘liq bo‘lib, og‘zaki nutq jarayonida ro‘y beradigan fonetik hodisalar morfemalarning sirtqi shakli, tuzilishi va qo‘llanish sohalariga sezilarli darajada ta‘sir qiladi. Bu holat ayniqsa nutqni avtomatik tahlil qilish, morfemalarni aniqlash va matnni tilshunoslik yondashuvida strukturaviy tahlil qilishda muhim ahamiyatga ega. Quyida ushbu fonetik hodisalar **kun.uz** saytida e‘lon qilingan materiallar asosida tahlil qilinadi:

1. Qisqarish (eliziya) - bu og'zaki nutqda tovushlarning tushib qolishi orqali so'z shaklining o'zgarishi hodisasi bo'lib, u ko'pincha nutq tezligining oshishi, talaffuz qulayligi yoki nutqiy stereotiplar ta'sirida yuzaga keladi. Eliziya natijasida ayrim morfemalar o'zining to'liq shaklida emas, balki qisqa yoki o'zgargan variantda namoyon bo'ladi. **Misol uchun:** “O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyev Samarqandga tashrif buyurdi” matni og'zaki nutqda ko'pincha **[Prezident Shavkat Mirziyoyev Samarqandga tashrif buyurdi]** tarzida aytiladi. Bu yerda “O'zbekiston Respublikasi” qisqargan, biroq mazmun o'zgarmagan. Bu morfologik jihatdan determinativ qismni fonetik jihatdan inkorporatsiyalash natijasidir[16].

2. Assimilatsiya — bu qo'shni tovushlarning o'zaro ta'siri natijasida bir tovush ikkinchisiga moslashib o'zgarishidir. Bu holat ayniqsa morfema chegaralarida ko'proq kuzatiladi. Fonetik o'zgarish morfemalarning aniq chegaralarini aniqlashda noaniqliklar tug'diradi. **Misol uchun:** “...Prezident Shavkat Mirziyoyev respublikamizda amalga oshirilayotgan islohotlar...” jumlasidagi “respublikamizda” so'zi tez aytilganda **[rezpublikamizda]** yoki **[rezbulikamizda]** tarzida talaffuz qilinishi mumkin, bu esa [s] tovushining [b] yoki [z] tovushlariga moslashganligini ko'rsatadi[17].

3. Koitasiya (so'zlar fonetik birikmasi) – bu so'zlar orasidagi fonetik chegaraning yo'qolib, bir butun ohangda aytilishi hodisasidir. Bu hodisa bog'lovchi morfemalarning nutqdagi ifodalanishiga bevosita ta'sir qiladi. **Misol uchun:** “O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyev Samarqandga tashrif buyurdi” jumlasini og'zaki nutqda **[O'zbekistonrespublikasiprezidentishavkatmirziyoyevsamarqandgatoshifbuyurdi]** tarzida uzluksiz aytilishi mumkin. Bu yerda morfemalar o'zining fonetik chegarasini yo'qotib, murakkab fonetik birlikni hosil qiladi[18].

4. Urg'u o'zgarishi (intonatsion siljish) – gapdagi yoki so'zdagi muhim qismlarni ajratib ko'rsatishga xizmat qiluvchi prosodik vosita hisoblanadi. Urg'uning o'zgarishi ayrim morfemalarning semantik yukini oshirish yoki kamaytirishga olib keladi. **Misol uchun:** “O'zbekiston Respublikasi Prezidenti **Shavkat Mirziyoyev** Samarqandga tashrif buyurdi” gapidagi urg'uning “Shavkat Mirziyoyev”ga tushishi bu shaxsga urg'u berilganligini ko'rsatadi. Ammo matn intonatsiyasida “Prezidenti” so'ziga urg'u berilsa, bu lavozimga urg'u bergan bo'ladi: “**O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyev...**”[16].

5. Epenteza (tovush qo'shilishi) — bu fonetik muvozanatni saqlash yoki talaffuzni qulaylashtirish maqsadida so'z yoki morfema tarkibiga ortiqcha tovushning kirib kelishi hodisasidir. Bu holat ayrim morfemalarning fonetik ko'rinishini o'zgartiradi, ayniqsa og'zaki nutqda aniqlik pasayishiga sabab bo'lishi mumkin. **Misol uchun:** “Ular yirik investitsiyalarni jalb qilishga erishmoqchi”

jumlasidagi “jalb” fe’li og‘zaki nutqda **[jalib]** tarzida talaffuz qilinadi. Bu yerda /i/ tovushining ortiqcha kiritilishi orqali fonetik yumshatish amalga oshgan va bu morfemaning sirtqi shakliga ta’sir qilgan[19].

6. Metateza (tovushlarning o‘rin almashuvi) — bu tovushlarning o‘zaro o‘rin almashtirishi natijasida yuzaga keladigan hodisadir. Og‘zaki nutqda tez va emotsional tarzda so‘zlashish natijasida bunday holatlar kuzatiladi va bu morfemalarning standart shaklidan og‘ishini bildiradi. **Misol uchun:** “Shartnomani buzgan tomon javobgarlikka tortiladi” jumlasidagi “shartnoma” so‘zi ayrim hollarda og‘zaki nutqda **[sharntoma]** tarzida talaffuz qilinadi. Bu tovushlar o‘rnining almashishi orqali morfemaning fonetik tuzilishi buziladi, bu esa morfologik tahlilda qiyinchilik tug‘diradi[20].

7. Reduksiya (tovushning zaiflashuvi yoki tushishi) — bu urg‘usiz holatda tovushlarning zaiflashuvi yoki butunlay tushib qolishidir. Bu hodisa morfemalarning to‘liq shaklini saqlab qolmasdan, qisqartirilgan shakllarda nutqda ishtirok etishiga olib keladi. **Misol uchun:** “Prezident tashabbusi bilan yangi loyiha ishga tushirildi” gapidagi “tashabbusi” so‘zi tez aytilganda **[tashabsi]** tarzida qisqarib aytilishi mumkin. Bu yerda [u] tovushining reduksiyaga uchragani morfemaning sirtqi shaklini buzadi[21].

8. Tilshunoslikda fonetik variantlarning morfologik neutralizatsiyasi. Ba’zi fonetik o‘zgarishlar natijasida bir necha morfema shakllari fonetik jihatdan bir hil eshitilishi mumkin. Bunday holatda morfologik ajratish qiyinlashadi. Bu tilshunoslikda **morfologik neutralizatsiya** deb ataladi. **Misol uchun:** “Yangi korxonalar tashkil etiladi” va “Yangi korxonani tashkil etadi” jumllaridagi “etiladi” va “etadi” fe’llari og‘zaki nutqda tez aytilganda **[etadi]** tarzida yangraydi, natijada zamon va nisbat farqlari fonetik jihatdan yo‘qoladi va morfologik ajratish imkonsizlashadi[22].

Yuqorida tahlil qilingan fonetik hodisalar – epenteza, metateza, reduksiya va morfologik neutralizatsiya – morfemalarning og‘zaki nutqdagi shakllanishiga va ulardan mazmuniy birlik sifatida foydalanish jarayoniga chuqur ta’sir ko‘rsatadi. Shu sababli, fonetik o‘zgarishlarning morfologiyaga ta’sirini aniqlash faqat nazariy tilshunoslik uchun emas, balki tabiiy tilni qayta ishlash, avtomatik morfologik tahlil, nutqni tanish va matnni kodlash kabi amaliy sohalar uchun ham muhim hisoblanadi.

O‘zbek tilida og‘zaki nutqda yuzaga keladigan fonetik o‘zgarishlar morfemalarning to‘liq yoki an’anaviy shaklini buzadi. Bu holat, ayniqsa, **korpus lingvistikasida** avtomatik morfologik tahlil tizimlari uchun muammo tug‘diradi. Masalan, **uzbekcorpus.uz** platformasida matnli ma’lumotlar yozma shaklda mavjud bo‘lsada, ushbu matnlarda aks ettirilgan og‘zaki nutq elementlari (yoki og‘zaki nutqqa xos bo‘lgan fonetik variantlar) tahlil qilinayotganda noto‘g‘ri morfologik ajratmalarga olib kelishi mumkin. Kun.uz saytida e’lon qilingan maqolada:

“Loyihaga jalb etilgan xorijiy investorlar...” Ushbu gap og‘zaki nutqda ko‘pincha quyidagi tarzda ifodalanadi: *[Loyihaga jalib qilgan xorijiy investorlar...]* Bu yerda “jalb” fe‘lining **epentezaga** uchrashi (ya’ni “jalib”) morfologik tahlil tizimlari uchun bu so‘zni yangi morfema deb qabul qilish ehtimolini kuchaytiradi. Bu esa morfologik xatoliklar sonini oshiradi.

Audiomatnlarda morfologik tahlil bosqichi nutq birliklarini grammatik jihatdan strukturaviy tahlil qilish, ularni yozma korpusga moslashtirish va avtomatik tizimlar uchun ma’lumotlar bazasi sifatida tayyorlash imkonini beradi. Ushbu bosqich, ayniqsa, tabiiy tilni avtomatik qayta ishlash (NLP) sohasida tilshunoslikka oid chuqur bilimlar bilan uyg‘un holda amalga oshirilganda, til birliklarini aniq identifikatsiya qilish, ularning grammatik shakllarini ajratish va tahlil qilishda muhim o‘rin egallaydi.

O‘zbek tilining agglutinativ xususiyati, ya’ni grammatik ma’no va vazifalarni ko‘rsatishda affikslarning izchilligi orqali amalga oshirilishi, morfologik tahlilni yanada murakkablashtiradi. Shu bilan birga, fonetik o‘zgaruvchanlik – ya’ni affikslar yoki asoslarning talaffuzdagi o‘zgarishi (masalan, tovush tushishi, assimitatsiya, dissimitatsiya) – avtomatlashtirilgan tahlil jarayonida hisobga olinishi kerak bo‘lgan muhim omildir. Aynan shu nuqtalarda morfologik tahlilning puxta ishlab chiqilgan modelga asoslanishi zarur bo‘ladi.

uzbekcorpus.uz platformasi bu borada amaliy tayanch nuqtasi bo‘lib xizmat qiladi. U yirik va xilma-xil matnlar asosida tuzilgan bo‘lib, so‘z shakllarining real qo‘llanilish namunalari bilan ta’minlaydi. Bu esa nafaqat morfologik tahlilning avtomatlashtirilgan modelini ishlab chiqishda, balki fonetik o‘zgarishlarning statistik tahlilini olib borishda ham katta imkoniyatlar yaratadi. Platformadagi ma’lumotlar morfologik tahlil bosqichining ishonchliligini oshirish bilan birga, algoritmlarni o‘rgatishda sifatli tayanch korpus sifatida xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova, N. (2021). O‘zbek tili elektron korpusining kompyuter modellari. Filologiya fanlari doktori (DSc) dissertatsiyasi avtoreferati.
2. Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. arXiv preprint arXiv:2006.11477.
3. Tursunov, J. (2023). "Audiomatnlar va zamon shakllari." O‘zbek tilshunosligi jurnali, №2, 45–53.
4. Mahmudov, N. (2015). O‘zbek tili morfologiyasi: affiksial tizim va uning amaliy tahlili. Toshkent: O‘zbekiston Milliy ensiklopediyasi. – 45b.
5. Cotterell, R., Kirov, C., Sylak-Glassman, J., & Eisner, J. (2015). Morphological segmentation for low-resource languages: Challenges and strategies.

Proceedings of the NAACL Workshop on Computational Morphology and Phonology.

6. Haspelmath, M., & Sims, A. D. (2010). Understanding Morphology (2nd ed.). London: Hodder Education. – 67b.

7. Johanson, L., & Csató, É. Á. (2021). Turkic Languages. 2nd edition. London: Routledge. DOI: <https://doi.org/10.4324/9780429025569> – 120b.

8. Mahmudov N., Nurmonov A. (2007). O'zbek tilining nazariy grammatikasi. Toshkent: Fan. — 134–137-bb.

9. Saidova Z. (2015). Hozirgi o'zbek tilining funksional grammatikasi. Toshkent: Akademnashr. — 93b.

10. Abdullayev F. (2021). O'zbek tilida affikslarning semantikasi. Toshkent: Fan. — 57b.

11. Mirtojiev M. (2010). O'zbek tilida grammatik shakllar va kategoriyalar. Toshkent: Universitet. — 112-bet.

12. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. Relevant content: Chapter 11 – Part-of-Speech Tagging. – 341-345b.

13. Nivre, J., de Marneffe, M. C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., ... & Zeman, D. (2016). Universal dependencies v1: A multilingual treebank collection. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 1659-1666.

14. Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed.). New Jersey: Prentice Hall. – 164-170b.

15. Habash, N. Y. (2010). Introduction to Arabic Natural Language Processing. Morgan & Claypool Publishers. – 61-66b.

16. “Shavkat Mirziyoyev Samarqandga tashrif buyurdi”, kun.uz, 2023-yil 12-oktabr.

17. “Prezident islohotlar haqida fikr bildirdi”, kun.uz, 2024-yil 1-mart.

18. “Prezident Shavkat Mirziyoyev yangi farmonni imzoladi”, kun.uz, 2023-yil 5-noyabr.

19. “Prezident investitsiyalarni jalb qilishga alohida e'tibor qaratmoqda”, kun.uz, 2024-yil 14-fevral.

20. “Shartnoma talablari buzilgan taqdirda...”, kun.uz, 2023-yil 9-iyun. [<https://kun.uz/News/2023/06/09/shartnoma>]

21. “Prezident tashabbusi bilan qator loyihalar amalga oshirildi”, kun.uz, 2023-yil 20-aprel. [<https://kun.uz/News/2023/04/20/tashabbus>]

22. “Yangi korxonalar tashkil etiladi”, kun.uz, 2023-yil 7-sentabr. [<https://kun.uz/News/2023/09/07/korxona>]

23. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In:



Abstract of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018).

24. Kubedinova L. Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101.

25. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.

26. N. Abdurakhmonova, U. Tuliyeu and A. Gatiatullin, "Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670043.