



UDC: 811.512.133'233

Zuxro AKBAROVA,

DSc, professor, Farg'ona davlat universiteti, Farg'ona, O'zbekiston
E-mail: akbarovazuhra41@gmail.com, ORCID: 0000-0002-8515-2868

Zarnigor UBAYDULLAYEVA,

Tadqiqotchi, Farg'ona davlat universiteti, Farg'ona, O'zbekiston
E-mail: zarnigorubaydullayeva@gmail.com, ORCID: 0000-0002-7300-3136

O'ZBEK BOLALAR NUTQINING LONGITUDINAL KORPUSINI SHAKLLANTIRISH: ILMIY ZARURAT VA CHILDES/TalkBank ASOSIDAGI YONDASHUV

Аннотация

Maqolada o'zbek bolalar nutqining longitudinal korpusini shakllantirishning ilmiy zarurati va CHILDES/TalkBank tajribasidan foydalanish yo'llari tahlil qilinadi. CHILDES korpuslari indeksida o'zbek tili alohida bolalar nutqi resursi sifatida qayd etilmagani, mavjud milliy tadqiqotlarning esa yagona longitudinal va qayta foydalaniladigan formatda birlashtirilmagani asosiy ilmiy bo'shliq sifatida baholanadi. Konseptual-tahliliy yondashuv asosida RusLan-M, Balausa1, turk tili materiallari, CHAT transkripsiyasi va o'zbek tili uchun Universal Dependencies resurslari qiyosiy o'rganildi. Natijada o'zbek korpusi uchun besh tayanch yo'nalish belgilandi: muntazam longitudinal kuzatuv, tabiiy audio-video material va metama'lumot, CHAT asosidagi transkripsiya, agglutinativ tuzilishga mos morfosintaktik belgilash, shuningdek, etik va texnik sifat nazorati. Taklif etilgan yondashuv erta leksik-grammatik rivojlanish, grammatik kategoriyalarning produktivlashuvi va individual nutqiy dinamikani korpus asosida o'rganish imkonini beradi. Korpusning CHILDES/TalkBankka kiritilishi esa material xalqaro texnik va etik talablarga javob berganidan keyingi bosqich sifatida qaraladi.

Kalit so'zlar: bolalar nutqi, longitudinal korpus, CHILDES, TalkBank, CHAT, korpus lingvistikasi, morfosintaktik belgilash, Balausa 1.

ФОРМИРОВАНИЕ ЛОНГИТУДНОГО КОРПУСА УЗБЕКСКОЙ ДЕТСКОЙ РЕЧИ: НАУЧНАЯ НЕОБХОДИМОСТЬ И ПОДХОД НА ОСНОВЕ CHILDES/TalkBank

Аннотация

В статье рассматривается научная необходимость формирования лонгитюдного корпуса узбекской детской речи и возможности использования опыта CHILDES/TalkBank. Основным исследовательским пробелом определяется отсутствие узбекского языка как самостоятельного корпуса детской речи в актуальном индексе CHILDES, а также единого лонгитюдного и повторно используемого формата для имеющихся национальных данных. На основе концептуально-аналитического подхода сопоставляются RusLan-M, Balausa1, материалы по турецкой детской речи, формат CHAT и ресурсы Universal Dependencies для узбекского языка. Выделены пять базовых направлений: регулярное лонгитюдное наблюдение, естественные аудио- и видеоданные с метаданными, транскрипция CHAT, морфосинтаксическая разметка с учётом агглютинативной структуры узбекского языка, этический и технический контроль качества. Предлагаемый подход создаёт основу для изучения раннего лексико-грамматического развития, продуктивного освоения грамматических категорий и индивидуальной динамики речи. Официальное включение корпуса в CHILDES/TalkBank рассматривается как последующий этап при выполнении требований репозитория.

Ключевые слова: детская речь, лонгитюдный корпус, CHILDES, TalkBank, CHAT, корпусная лингвистика, морфосинтаксическая разметка, Balausa 1.

DEVELOPING A LONGITUDINAL CORPUS OF UZBEK CHILD SPEECH: SCHOLARLY NEED AND A CHILDES/TalkBank-BASED APPROACH

Annotation

This article examines the scholarly need for a longitudinal corpus of Uzbek child speech and considers how the CHILDES/TalkBank framework can inform its development. The absence of Uzbek as a dedicated child-language corpus in the current CHILDES index, together with the lack of a unified longitudinal and reusable format for existing national data, is identified as the central research gap. Using a conceptual-analytical approach, the study compares RusLan-M, Balausa1, Turkish child-language data, CHAT transcription practices, and Universal Dependencies resources for Uzbek. Five core requirements are proposed: regular longitudinal sampling, naturalistic audio-video data with metadata, CHAT-based transcription, morphosyntactic representation adapted to Uzbek agglutinative structure, and ethical and technical quality control. The framework is intended to support corpus-based study of early lexical-grammatical development, the productive emergence of grammatical categories, and individual developmental dynamics. Formal contribution to CHILDES/TalkBank is treated as a subsequent stage contingent on meeting repository requirements.

Keywords: child language, longitudinal corpus, CHILDES, TalkBank, CHAT, corpus linguistics, morphosyntactic annotation, Balausa 1.

Kirish. Bolalar nutqini o'rganish til ontogenezining universal va tilga xos qonuniyatlarini aniqlashda muhimdir.

Ayniqsa longitudinal yondashuv bir bola yoki bir guruh bolalarning nutqiy rivojlanishini vaqt davomida muntazam

kuzatish imkonini beradi [Berman and Slobin 1994; MacWhinney 2000].

Raqamli korpus infratuzilmalari bunday materialni tekshiriladigan va qayta foydalaniladigan ilmiy resursga aylantiradi. CHILDES (Child Language Data Exchange System) TalkBank tizimining bolalar tiliga oid qismi bo'lib, CHAT transkripsiyasi, media, metama'lumot va CLAN tahlil vositalarini o'zaro bog'laydi [MacWhinney 2019; MacWhinney and Fromm 2022].

O'zbek tilshunosligida bolalar nutqiga oid tadqiqotlar muayyan ilmiy zamin yaratgan [Qurbonova 2009; Masharipova and Qutlimuratova 2024; Abduraxmonova 2025; Hamidova 2026]. Biroq ular turli maqsad, yosh va formatlarda shakllangan; CHILDES indeksida esa o'zbek tili alohida bolalar nutqi korpusi sifatida qayd etilmagan [TalkBank 2026a]. Shu bois yagona longitudinal, media bilan bog'langan va xalqaro standartlarga mos resursga ehtiyoj saqlanmoqda.

Maqoladan ko'zda tutilgan maqsad, o'zbek bolalar nutqining longitudinal korpusini shakllantirish zaruratini ilmiy jihatdan asoslash va CHILDES/TalkBank tajribasiga tayangan holda uning tayanch metodologik komponentlarini belgilashdir. Tadqiqotning yangiligi o'zbek bolalar nutqi uchun korpusning besh o'zaro bog'liq komponentdan iborat konseptual modelini taklif qilishda ko'rinadi: longitudinal namuna olish; tabiiy media va metama'lumot; CHAT transkripsiyasi; o'zbek tilining agglutinatuv tuzilishiga mos morfosintaktik belgilash; etik-texnik sifat nazorati va xalqaro taqdimotga tayyorgarlik.

Mazkur maqola konseptual-tahliliy xarakterga ega bo'lib, tahlil uchun bolalar nutqi korpuslari, CHILDES/TalkBank metodologiyasi, CHAT transkripsiyasi, Universal Dependencies hamda o'zbek bolalar nutqiga oid milliy tadqiqotlar tanlandi. Tadqiqot uchun manbalarni saralashda uch mezon asos qilib olindi: tabiiy bolalar nutqi va longitudinal kuzatuvga daxldorlik; transkripsiya, media va lingvistik belgilash bilan bog'liqlik; o'zbek tilining agglutinatuv xususiyatlari uchun metodologik ahamiyat.

Qiyosiy-tipologik va korpus-analitik tahlilda RusLan-M, Balausa1 va turk tili bo'yicha CHILDES materiallari metodologik analog sifatida ko'rib chiqildi; o'zbek morfosintaktik belgilashining texnik tayanchi sifatida Uzbek-UT UD treebanki tahlil qilindi [Lelik et al. 2025; TalkBank 2026b; Akhundjanova and Talamo 2025].

1-jadval. CHILDES korpuslari tajribasi va o'zbek korpusini shakllantirish uchun muhim mezonlar

Korpus / ma'lumotlar to'plami	Til va tadqiqot turi	Audio-video materiallar	Lingvistik belgilash	O'zbek korpusini shakllantirishda e'tiborga olinadigan jihatlar
RusLan-M	Rus tili; longitudinal tadqiqot; 2 bola	Video va audio	CHAT/CHILDES asosida transkripsiya; shaxsiy ma'lumotlar himoyalangan	Bir xil bolalarni uzoq muddat davomida muntazam kuzatish; shaxsiy ma'lumotlar maxfiylikni ta'minlash; audio-video materiallar bilan transkriptlarning o'zaro bog'liqligini saqlash
Balausa1	Qozoq tili; asosan ko'ndalang kesimli tadqiqot; 8 bola	Audio; taxminan 5 soat	%mor qatlami Batchalign2 va UD asosida yaratilgan hamda ona tili sohiblari tomonidan tekshirilgan	Agglutinatuv turkiy til materialida Universal Dependencies imkoniyatlarini qo'lda tekshirish bilan uyg'unlashtirish tajribasidan foydalanish
Aksu Turkish	Turk tili; bolalarning tabiiy nutqi	Tarixiy nutqiy ma'lumotlar to'plami	CHILDES transkriptlari	Turkiy tillarda morfologik kategoriyalarning egallanish jarayonini qiyosiy o'rganish uchun metodologik asos yaratishi

Mavjud xalqaro tajriba va o'zbek tilining lingvistik xususiyatlarini umumlashtirish asosida o'zbek bolalar nutqi korpusining beshta asosiy tarkibiy yo'nalishi belgilandi. Birinchi komponent - muntazam longitudinal namuna: bolani faol va tabiiy vaziyatlarda qayta-qayta kuzatish, har bir kuzatish jarayoni oralig'ini imkon qadar barqaror saqlash va uzilishlarni alohida qayd etib borish [TalkBank 2026d]. Ikkinchi

Longitudinal korpusning asosiy ilmiy qiymati shundaki, u asosan natijani emas, rivojlanish jarayonini hujjatlashtiradi. Bolaning tildagi ma'lum bir affiksni ilk bor qo'llashi hali uni egallaganini anglatmaydi: ushbu shakl keyingi kuzatuvlarda takroran namoyon bo'lmasligi, dastlab ayrim leksemalar doirasida qo'llanishi, keyinchalik esa boshqa leksemalarga ham kengayib, produktiv xususiyat kasb etishi mumkin. Shu sababli morfologik kategoriya ontogenezi "bor/yo'q" ko'rinishida emas, paydo bo'lish – takror qo'llanish – umumlashish – produktivlashish ketma-ketligida kuzatish zarur.

Mazkur masala o'zbek tili uchun alohida ilmiy ahamiyat kasb etadi. O'zbek tilida otlarda ko'plik, egalik va kelishik, fe'llarda esa inkor, zamon, mayl va shaxs-son ma'nolari affiksalar orqali ifodalanadi. Masalan, kitob-ni, uy-im-ga, bor-di-m kabi shakllarda bola bir so'z doirasida bir nechta grammatik munosabatni o'zlashtirishi kerak. Longitudinal material ushbu affikslarning qaysi tartibda paydo bo'lishi, qaysi shakllar dastlab leksik qolip sifatida qo'llanishi va qachondan turli so'zlarga erkin birikishi kabi savollarni empirik tekshirish imkonini beradi. Bu yondashuv turk tilini o'zlashtirish bo'yicha tadqiqotlarda aniqlangan morfologik rivojlanish qonuniyatlarini o'zbek tili materiali bilan qiyoslash imkonini beradi [Aksu-Koç and Slobin 1985].

Longitudinal tadqiqot dizayni individual rivojlanish farqlarini kuzatish imkonini beradi. Bunda guruh o'rtacha ko'rsatkichlari ortida turli rivojlanish yo'llari yashirinishi mumkin. Bola va unga yaqin kattalar o'rtasidagi tabiiy muloqotni bir kuzatuv jarayonida qayd etish kattalar nutqiy ta'siri bilan bolaning keyingi nutqiy rivoji o'rtasidagi bog'liqlikni tahlil qilish imkonini beradi [Newman, Rowe, and Bernstein Ratner 2016].

CHILDES korpuslari tadqiqot maqsadiga ko'ra turlicha tuzilgan bo'lsa-da, ularda ishtirokchilarni aniq belgilash, media va transkriptni bog'lash, yagona transkripsiya qoidalariga rioya qilish, metama'lumotlarni tizimli qayd etish hamda korpus yaratish jarayonini hujjatlashtirish umumiy metodologik talablar sifatida namoyon bo'ladi. Quydagi jadval o'zbek loyihasi uchun metodologik jihatdan yaqin uch tajribani umumlashtiradi.

komponent – tabiiy audio-video materialni metama'lumot bilan birga yuritish. Har bir qayd, bola kodi, aniq yoshi, sana, kommunikativ vaziyat, suhbatdoshlar va til muhiti haqidagi ma'lumotlar bilan bog'lanishi lozim.

Uchinchi komponent - CHAT asosidagi transkripsiya va media bilan vaqt bo'yicha moslashtirish. CHATning vazifasi bolalar nutqini adabiy me'yorga moslashtirish emas, balki

bolaning amalda aytgan nutqini keyinchalik qayta tahlil qilish mumkin bo'lgan shaklda saqlashdir. Tovush almashtirish, qisqartirish, takror, tugallanmagan birlik va ustma-ust gapirish kabi hodisalar loyiha qo'llanmasida bir xil qarorlar asosida belgilanishi kerak [TalkBank 2026c].

To'rtinchi komponent - o'zbek tilining agglutinatив tuzilishiga mos morfosintaktik belgilash. Uzbek-UT treebanki lemma, so'z turkumi, morfologik xususiyat va gap tarkibidagi sintaktik bog'lanishlarini taqdim etadi [Akhundjanova and Talamo 2025], biroq kattalar yozma matniga mos modelni spontan bola nutqiga bevosita ko'chirish mumkin emas. Fonetik reduksiya, affiks tushishi, ellipsis va tugallanmagan birliklar sabab avtomatik tahlil ekspert tekshiruvi bilan to'ldirilishi lozim. Korpusning asosiy qismida morfosintaktik qatlam, tanlangan reprezentativ qismida esa chuqurroq fonetik-pragmatik belgilash qo'llanishi mumkin.

Beshinchi komponent - etika va sifat nazorati. Bola ovozi va videosi shaxsni aniqlash imkonini beruvchi ma'lumot bo'lishi mumkinligi sababli, ota-ona yoki qonuniy vakil roziligi, shaxsiy ma'lumotlarni shartli kodlar bilan almashtirish va ma'lumotlardan foydalanish darajalari loyihaning dastlabki bosqichidayoq belgilanadi. CHILDES ayrim korpuslarda transkript va video materiallardan foydalanish darajalarini turlicha belgilaydi [TalkBank 2026e]. Texnik sifat fayl nomlari, ishtirokchi va media haqidagi ma'lumotlar, media bilan transkriptning o'zaro mosligi hamda CHAT formatining tekshiruvi orqali nazorat qilinadi.

Sintaktik qatlamda bazaviy SOV tartibi (Subject-Object-Verb, ya'ni ega-to'ldiruvchi-kesim), komponentlarning tushirilishi, obyektning kelishik bilan ifodalanishi, egalik-qaratqich munosabatlari va gaplarning murakkablashuvi tahlil qilinishi mumkin. Bu ma'lumotlar o'zbek tilini turk va qozoq kabi qarindosh tillar hamda rus tili kabi tipologik jihatdan farqli til bilan qiyoslashga zamin yaratadi.

Taklif etilayotgan model yakuniy korpus natijasi emas. Uni pilot longitudinal materiallarda sinash, CHAT fayllari va annotatsiyani ekspertlar tomonidan tekshirtirish hamda TalkBank talablariga muvofiqligini baholash zarur. CHILDES/TalkBankka integratsiya haqida faqat rasmiy qabul tasdiqlangandan keyin xulosa qilish mumkin.

Xulosa qilib, o'zbek bolalar nutqining longitudinal korpusi leksik-grammatik rivojlanish va individual nutqiy dinamikani vaqt bo'yicha kuzatish uchun tizimli empirik asos yaratadi. Taklif etilgan metodologik model muntazam longitudinal kuzatuv, tabiiy audio-video materiallar va metama'lumotlar, CHAT transkripsiyasi, o'zbek tilining agglutinatив xususiyatlariga mos morfosintaktik belgilash hamda etik-texnik sifat nazoratini birlashtiradi. Keyingi bosqichda pilot materiallarni CHAT formatida transkripsiyalash, morfosintaktik belgilashni ekspert tekshiruvidan o'tkazish va topshiruv hujjatlarini TalkBank talablariga moslashtirish lozim.

ADABIYOTLAR

1. Abduraxmonova, Sohba Aktam qizi. 2025. "O'zbek bolalar nutqining korpusga asoslangan psixolingvistik tadqiqi: CHILDES tizimi asosida yondashuv." Contemporary Technologies of Computational Linguistics – CTCL.2025, 538–540.
2. Akhundjanova, Arofat, and Luigi Talamo. 2025. "Universal Dependencies Treebank for Uzbek." Proceedings of the Third Workshop on Resources and Representations for Under-Resourced Languages and Domains, 1–6.
3. Aksu-Koc, Ayhan, and Dan I. Slobin. 1985. "The Acquisition of Turkish." In The Cross-Linguistic Study of Language Acquisition, Vol. 1, 839–878. Hillsdale, NJ: Lawrence Erlbaum Associates.
4. Berman, Ruth A., and Dan I. Slobin, eds. 1994. Relating Events in Narrative: A Crosslinguistic Developmental Study. Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Hamidova, Gulrux Axmat qizi. 2026. "Maktab yoshidagi o'zbekzabon bolalar og'zaki nutqini MAIN metodologiyasi asosida o'rganish." Tadqiqotlar 88 (6): 110–112.
6. Lelik, Valeriia, Mariia Diachkova, Svetlana V. Dorofeeva, Anastasiya Lopukhina, Olga Dragoy, and Irina A. Sekerina. 2025. RusLan-M Corpus. CHILDES/TalkBank. <https://doi.org/10.21415/1XMG-D508>.
7. MacWhinney, Brian. 2000. The CHILDES Project: Tools for Analyzing Talk. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.
8. MacWhinney, Brian. 2019. "Understanding Spoken Language through TalkBank." Behavior Research Methods 51: 1919–1927. <https://doi.org/10.3758/s13428-018-1174-9>.
9. MacWhinney, Brian, and Davida Fromm. 2022. "Language Sample Analysis With TalkBank: An Update and Review." Frontiers in Communication 7: 865498. <https://doi.org/10.3389/fcomm.2022.865498>.
10. Masharipova, Baxtigul, and Barno Qutlimuratova. 2024. "O'zbek tili bolalar nutqi audio korpusini yaratish." Computer Linguistics: Problems, Solutions, Prospects 1 (1).
11. Newman, Rochelle S., Meredith L. Rowe, and Nan Bernstein Ratner. 2016. "Input and Uptake at 7 Months Predicts Toddler Vocabulary." Journal of Child Language 43 (5): 1158–1173.
12. Qurbonova, Munavvara A. 2009. O'zbek bolalar nutqi leksikasining sotsiopsixolingvistik tadqiqi. Filologiya fanlari nomzodi dissertatsiyasi. Toshkent: O'zbekiston Milliy universiteti.
13. TalkBank. 2026a. "CHILDES Corpora." <https://talkbank.org/childes/access/index.html> (murojaat sanasi: 26.08.2026).
14. TalkBank. 2026b. "CHILDES Kazakh Balasaul Corpus." <https://talkbank.org/childes/access/Other/Kazakh/Balasaul.html> (murojaat sanasi: 26.08.2026).
15. TalkBank. 2026c. "CHAT Transcription Format." <https://talkbank.org/0info/manuals/CHAT.html> (murojaat sanasi: 26.08.2026).
16. TalkBank. 2026d. "Creating a Child Language Corpus." <https://talkbank.org/childes/topics/newcorpus.html> (murojaat sanasi: 26.08.2026).
17. TalkBank. 2026e. "CHILDES Data Access Levels." <https://talkbank.org/childes/access.html> (murojaat sanasi: 26.08.2026).