

## CHILDES KORPUSLAR TIZIMIDA O‘ZBEK TILI UCHUN MA’LUMOTLAR BAZASI YARATISH

**O‘rinboyeva Ro‘za Toliboyevna**

Mirzo Ulug‘bek nomidagi

O‘zbekiston Milliy universiteti magistranti.

E-mail: [orinboyevaroza@gmail.com](mailto:orinboyevaroza@gmail.com)

**Allaberdiyeva Durdona Gurbanmurat qizi**

Mirzo Ulug‘bek nomidagi

O‘zbekiston Milliy universiteti magistranti.

E-mail: [durdonallaberdiyeva@gmail.com](mailto:durdonallaberdiyeva@gmail.com)

**Annotatsiya:** Bolalar nutqini tahlil qilish tilshunoslikdagi murakkab vazifalardan biri hisoblanib, jahon korpusshunosligida bu masalaga oid tajribalar CHILDES tizimida jamlangan. Ushbu maqolada o‘zbek tilidagi bolalar nutqini o‘rganishga qaratilgan tadqiqotlar asosida CHILDES tizimi uchun ma’lumotlar bazasi yaratish masalasi yoritib berilgan.

**Kalit so‘zlar:** transkript, audio, video, klinik nutq, Praat, Phon.

**Abstract:** Analysis of children’s speech is considered one of the most complex tasks in linguistics, and experiences related to this issue in world corpus studies are gathered in the CHILDES system. This article describes the issue of creating a database for the CHILDES system based on research aimed at studying children’s speech in the Uzbek language.

**Keywords:** transcript, audio, video, clinical talk, Praat, Phon.

**Аннотация:** Анализ детской речи считается одной из самых сложных задач в лингвистике, и опыт, связанный с этим вопросом в мировых корпусных исследованиях, собран в системе CHILDES. В данной статье описан вопрос создания базы данных для системы CHILDES на основе исследований, направленных на изучение речи детей на узбекском языке.

**Ключевые слова:** стенограмма, аудио, видео, клинический разговор, Praat, Фон.

### Kirish

Dunyodagi jamiki tillar o‘zining fonetik, fonologik, sintaktik, semantik, morfologik kabi xususiyatlariga ega bo‘lib, tilshunoslikda ular ma’lum qonun-qoidalarga solingan. Nutqning kishilik jamiyatidagi barcha ko‘rinishi muayyan tadqiqot ishlari uchun obyekt vazifasini o‘tamoqda. Ular orasida bolalar nutqini o‘rganish, ularning transkripsiyasini yaratish ham qiziqarli, ham mashaqqatli vazifalardan biridir.

CHILDES – bola tilidagi ma’lumotlar almashinuvi tizimi bo‘lib, bolalar nutqini o‘rganishga qaratilgan korpuslar jamlanmasi hisoblanadi. Olimlar bolalarda uchraydigan afaziya, ikkinchi tilni o‘zlashtirish hamda nutqni tahlil qilishda yuzaga keladigan muammolarga yechim topish maqsadida turli xil tajribalarni amalga oshirishgan. Ularning nutqi, har bir so‘zi audio, video shaklida yozib olingan. Muayyan darajada transkripsiyada ifodalangan. Ammo CHILDES


tarkibiga o'zbek tilidagi bolalar nutqini o'rganishga qaratilgan tadqiqotlar o'rin olmagan. Ushbu maqolada bu masalaning dolzarbligi va uning amaliy ahamiyati yoritib berilgan.

### Adabiyotlar tahlili va metodlar

1984-yilda Karnegi Mellon universitetining psixologiya bo'limida doktor Brayan MakVinni va doktor Ketrin Snou tomonidan tashkil etilgan CHILDES dasturi bolalar tilini o'zlashtirish transkriptlari va yozuvlarini tadqiqotchilarga bepul, ommaviy ma'lumotlar to'plami sifatida taqdim etishni maqsad qildi [MacWhinney., Snow K., 1985]. Mualliflar bu boradagi yuzaga keladigan muammolarni bartaraf etishga kirishdi. Dasturning eng dastlabki transkriptlari 1960-yillarda taqdim etilgan. 1990-yillarning boshida CHILDES dan foydalanish qulayligi sababli bolalar tilini o'zlashtirish bo'yicha tadqiqotlar soni sezilarli darajada oshgan [Alessandro S., Stephan C. M., 2019]. 2015-yilga kelib CHILDES tarkibidagi 230ta korpus 26 tildagi transkriptlar, audio va video materiallarni o'zida jamladi. Bu butun dunyo bo'ylab bir vaqtning o'zida bir necha tillar kesimida bolalar nutqini o'rganish imkonini beradi.

CHILDESning bosh menyusi TalkBankdan iborat. Bunda o'rganilgan dunyo tillari 24 kategoriyaga ajratilgan. Ularning har biri korpusning yaratilish tarixi, muallifi haqida qisqacha ma'lumot, o'rganilgan bolalarning yoshi, soni, media ma'lumotning turi, izohi va qo'shimcha ma'lumotlar berilgan. CHILDESning asosiy ma'lumotlar bazasi PhonBankda joylashgan:


TalkBank







PhonBank Project

PhonBank is the child phonology component of the [TalkBank](#) system. TalkBank is a system for sharing and studying conversational interactions. PhonBank is supported by grant RO1-HD051698 from NIH-NICHHD to Brian MacWhinney and Yvan Rose. *PHON* is designed and built by Yvan Rose and Greg Hedlund. Currently available materials include:

System	Database	Phon Program
<a href="#">**Ground Rules**</a> <a href="#">Contributing New Data</a> <a href="#">IRB Principles</a> <a href="#">Hints on downloading</a> <a href="#">Database versioning</a>	<a href="#">**Index to Corpora**</a> <a href="#">Browsable Database</a> <a href="#">Derived Corpora</a> <a href="#">TalkBankDB database search</a>	<a href="#">Phon website and User Manual</a> <a href="#">Phon on GitHub</a> <a href="#">PhonTalk (CHAT ⇌ Phon)</a>
Links	Resources	Contact
<a href="#">Other TalkBank databases</a> <a href="#">Other Child Language sites</a> <a href="#">Research based on Phon</a> <a href="#">Workshop presentations</a>	<a href="#">Phon Basics: A practical introduction</a> <a href="#">Video tutorials on YouTube</a> Downloadable files for video tutorials: <ul style="list-style-type: none"> <li>• <a href="#">Phon projects for tutorials</a></li> <li>• <a href="#">Media files for tutorials</a></li> </ul>	Yvan Rose: <a href="#">homepage</a> <a href="#">Phon mailing list</a>



Legal Information

GDPR

Visits since August 28, 2017: 54,300

Bu sahifadagi "System" qismida korpus haqidagi asosiy qoidalar, foydalanish bo'yicha tavsiyalar, versiyalari haqida ma'lumotlar berilgan. "Database" qismida mavjud korpuslarga hissa qo'shgan olimlar va ularning ishlari ko'rsatilgan. "Phon Program" qismida mualliflar bilan bog'lanish uchun havolalar, minnatdorchilik xatlari hamda ma'lumotlarning har xil turdagi qurilmalar uchun moslangan shakllari o'rin olgan. "Links" asosiy qismlardan biri bo'lib, unda TalkBankning, bola tilini o'rganayotgan tadqiqot institutlari hamda olimlarning internet saytlariga havolalar keltirilgan. "Resources"da turli qo'llanmalar, You Tubedagi video darsliklar va media

fayllar ko'rsatilgan. So'nggi qismda CHILDES va TalkBank muammolari bilan bog'liq yettita Google guruhlarining elektron pochталari berilgan.

### **Tadqiqotning amaliy ahamiyati**

CHILDES korpuslar tizimida o'zbek tili uchun bolalar nutqiga asoslangan ma'lumotlar bazasini yaratish orqali korpus lingvistikasining taraqqiyotiga hissa qo'shish mumkin. Bu masalani hal etishda bir necha dasturiy ta'minotlardan foydalaniladi.

Dastlab bolalar nutqi bilan ishlashda Phondan foydalanamiz. Phon – bu matn va fonologik ma'lumotlar korpusini yaratishni qo'llab-quvvatlaydigan dasturiy ta'minot. U bolalar tilini rivojlantirishga qaratilgan bo'lsa-da, Phondan bolalar yoki kattalar tili ma'lumotlariga asoslangan korpus tadqiqotlarining deyarli barcha turlarida (masalan, matn, fonologik, akustik, klinik) foydalanish mumkin. Shuningdek, Phon fonologik birliklar (masalan, fonologik xususiyatlar, stress, ohanglar) bo'yicha tadqiqotlar uchun maxsus yordamni taklif qiladi va klinik nutqni tahlil qilish uchun ayrim maxsus funksiyalarni o'z ichiga oladi [Rose Y., Carol S.G., 2016].

Korpusshunoslikda nutqni tahlil qilish uchun, asosan, Praatdan foydalaniladi. Praat – bu nutq va boshqa audio signallarni tahlil, sintez va manipulyatsiya qilish uchun ishlatiladigan bepul dasturiy ta'minot. U Amsterdam universitetida Pol Boersma va Devid Venink tomonidan ishlab chiqilgan va tilshunoslik, fonetika hamda nutq patologiyasi kabi sohalarda tadqiqotchilar tomonidan keng qo'llaniladi. Praatning asosiy xususiyatlari quyidagilar:

1. Ovoz yozish va tahrirlash: Praat foydalanuvchilarga audio fayllarni yozib olish va tahrirlash imkonini beradi, jumladan, tovush to'liqlinlari va spektrogrammalarni ko'rish imkoniyati.

2. Nutqni tahlil qilish: Praat nutq signallarini tahlil qilish uchun keng ko'lamli vositalarni taqdim etadi: tovush balandligi, formantlar, intensivlik va boshqa akustik parametrlarni o'lchash.

3. Matnni nutqqa sintez qilish: Praat matn kiritishdan nutqni sintez qilish uchun ishlatilishi mumkin, bu foydalanuvchilarga tadqiqot maqsadlari uchun maxsus nutq stimullarini yaratishga imkon beradi.

4. Skript tili: Praat o'rnatilgan skript tiliga ega bo'lib, u foydalanuvchilarga vazifalarni avtomatlashtirish va maxsus tahlillarni yaratish imkonini beradi.

Tadqiqot jarayonida to'plangan audio va video materiallarning transkriptlari yaratiladi. Bolalardagi til o'rganish qobiliyati yosh darajasiga ko'ra tahlil qilinadi, ulardagi muammo va kamchiliklar o'rganilib, jahon tajribalari bilan taqqoslanadi. O'zbek tilidagi bolalar o'rganishi uchun qiyin bo'lgan xususiyatlar aniqlanadi. Ayni shu jihatlar korpusshunoslikning til taraqqiyotidagi ahamiyatini belgilab beradi [Abdurakhmonova, N. 2019, 2021, 2022].

### **Xulosa**

CHILDES tizimida o'zbek tili uchun korpus yaratish orqali bolalar nutqini o'rganish borasidagi tadqiqotlarning amaliy natijasini taqdim etish mumkin. Bu kelgusi tadqiqotlar uchun asos bo'la oladi. Bolalar nutqida uchraydigan nuqsonlarning oldini olishga, turli yoshdagi bolalarda ikkinchi tilni o'zlashtirishga doir muammalarning bartaraf etilishiga yordam beradi.

### **Foydalanilgan adabiyotlar:**

1. Rose Y., Carol S.G. Using PhonBank and Phon in Studies of Phonological Development and Disorders, 2016. Using PhonBank and Phon in Studies of Phonological Development and Disorders - PMC (nih.gov)

2. MacWhinney B. Morphosyntactic Analysis of the CHILDES and TalkBank Corpora. poster2 (psu.edu)
3. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
4. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O 'zbek tili elektron korpusida (<http://uzbekcorpus.uz/>) og 'zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. Academic research in educational sciences, 3(3), 644-650.
5. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
6. Sanchez A., Stephan M.C., Mika B., MacDonald K.E., Daniel Y., Michael C.F. Childes-db: A Flexible and reproducible interface to the child language data exchange system, 2019.childes-db: A Flexible and reproducible interface to the child language data exchange system
7. MacWhinney B., Snow C. The child language data exchange system//Journal of Child Language: 1985.
8. CHILDES Corpora (talkbank.org)
9. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
10. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).