

BOLALAR NUTQI KORPUSI SHAKLLANISH BOSQICHLARI

Normamatova Mashhura

O‘zbekiston Milliy Universiteti 2-bosqich talabasi.

E-mail: mashhuranormamatova0@gmail.com

Ilmiy rahbar: **Abduraxmonova Nilufar**

Mirzo Ulug‘bek nomidagi O‘zbekiston milliy universiteti

Professori, filologiya fanlari doktori.

E-mail: n.abduraxmonova@nuu.uz

Annotatsiya: Ushbu maqolada bolalar nutqi korpuslarining mohiyati hamda bolalar nutqi korpuslarining shakllanish bosqichlari borasida fikr yuritiladi. Bugungidek jadal rivojlanib borayotgan dunyoda korpus har qanday sohada muhim ahamiyat kasb etib borayotgan universal manbadir. Korpuslar rivojlanish tarixiga nazar tashlaydigan bo‘lsak, dastlabki korpuslar tilshunoslik sohasida vujudga kelgan. Bugungi kunda esa, korpuslar nafaqat tilshunoslikda balki, tibbiyot, pedagogika, psixologiya yo‘nalishlarida ham o‘z o‘rmini egallashda davom etmoqda. Shuningdek, korpuslar til o‘qitish sohasida ham muhim ahamiyatga molik hodisa bo‘ldi. Bu borada bolalar nutqi korpuslari, ayniqsa, katta rol o‘ynamoqda. Maqolada bolalar nutqi korpuslaridan ayrimlariga qisqacha to‘xtalib o‘tildi. Keltirilgan korpuslar korpus lingvistikasining bolalar nutqi korpusi yo‘nalishidagi eng muvaffaqiyatli ishlar sirasiga kiradi.

Kalit so‘zlar: bolalar nutqi korpusi, nutqni avtomatik aniqlash, CHILDES, Brayan MakVinni, TaklBank, Demut Sesoto korpusi, Ketrin Demuth.

Abstract: This article examines the nature of children's speech corpora and the stages of formation of children's speech corpora. In today's rapidly developing world, the body is a universal resource that is becoming increasingly important in any field. If you look at the history of the development of corpora, the first corpora were created in the field of linguistics. Today, corpora continue to occupy their place not only in linguistics, but also in medicine, pedagogy and psychology. In addition, corpora have become an important phenomenon in the field of language teaching. Children's speech corpora play a particularly important role in this regard. The article briefly touches on some corpora of children's speech. The given corpora are among the most successful works of corpus linguistics in the field of corpora of children's speech.

Keywords: child speech corpus, automatic speech recognition, CHILDES, Brian McWhinney, TalkBank, Demuth Sesotho Corpus, Catherine Demuth.

Аннотация: В данной статье рассматривается природа детских речевых корпусов и этапы формирования детских речевых корпусов. В современном быстро развивающемся мире корпус является универсальным ресурсом, приобретающим все большее значение в любой сфере. Если взглянуть на историю развития корпусов, то первые корпуса были созданы в области лингвистики. Сегодня корпуса продолжают занимать свое место не только в лингвистике, но и в медицине, педагогике и психологии. Кроме того, корпуса стали важным явлением в сфере преподавания языка. Особенно важную роль в этом отношении играют детские речевые корпуса. В статье кратко затронуты некоторые корпуса детской

речи. Приведенные корпуса относятся к числу наиболее успешных работ корпусной лингвистики в области корпусов детской речи.

Ключевые слова: корпус детской речи, автоматическое распознавание речи, CHILDES, Брайан Маквинни, TalkBank, Demuth Sesotho Corpus, Кэтрин Демут,

So‘ngi bir necha o‘n yilliklar davomida korpus lingvistikasi doirasida samarali ishlar qilindi. Bu davr mobaynida olimlar tomonidan turli korpuslar ishlab chiqildi. Biroq, korpus lingvistikasi borasidagi ko‘pgina tadqiqotlar kattalar nutqiga asoslangan edi. Bolalar nutqi kattalar nutqiga nisbatan artikulyar, akustik, jismoniy va lingvistik xususiyatlariga ko‘ra murakkab ekanligi bilan ancha murakkab jarayonlarni taqozo etardi. Ehtimol shu bois, bir muncha vaqt bolalar nutqi bilan ishlash e‘tibordan chetda qoldi. Bu esa ushbu soha tadqiqot ishlarining eng e‘tiborli nuqtasi ekanini anglatadi.

Nutq korpuslari borasidagi ishlarni o‘rganish jarayonida tadqiqotchida “bolalar nutqi korpusining mohiyati nimadan iborat?” degan savol tug‘ilishi tabiiy. Dastlab ushbu savolga atroflicha javob berishga harakat qilamiz.

Bir vaqtlar elektron ta‘lim kelajak ta‘limida nuhum rol o‘ynaydi deb qaralgan bo‘lsa-da, u an’anaviy ta‘lim tizimida tez sur‘atlarda o‘zgartirib yuborishi kutilmagan edi. COVID-19 pandemiyasi natijasida onlayn o‘qitish dasturlari, til ilovalari, video konferensiyalar, virtual repetitorlik muhim o‘qitish vositalariga aylandi. Barcha o‘rganuvchilar, jumladan, eshitish imkoniyati cheklangan bo‘lgan o‘rganuvchilar uchun real vaqt rejimida taglavhalar (subtitr)ni taqdim etish uchun nutqni aniqlash texnologiyalariga yanada ehtiyoj kuchaydi. Nutqni aniqlash texnologiyasi qo‘l yozuvi yoki imlo bilan bog‘liq qiyinchiliklarga duch kelgan bolalar uchun ham qulaylik yaratadi. Disleksiya yoki boshqa o‘rganishda nuqsonlari bo‘lgan bolalar qo‘lda yozish o‘rniga, o‘z ovozlari bilan foydalanib yozish uchun nutqdan matnga texnologiyasidan foydalanishi mumkin. Bolalar – multimedia o‘yinlari, o‘quv dasturlari va o‘quv materiallari nuqtai nazaridan avtomatik nutqni aniqlash texnologiyalaridan eng samarali foydalanadigan qatlamdir. Mana shu jihatlarni inobatga olgan holda bolalar nutqi korpusi korpus lingvistikasining muhim yo‘nalishlaridan biri desak mubolag‘a bo‘lmaydi.

Korpus lingvistikasi tarixiga nazar tashlaydigan bo‘lsak, bolalar nutqi korpusi mavzusi nisbatan yangi masalalardan hisoblanadi. Shunga qaramasdan, nutqni avtomatik aniqlash texnologiyasiga asoslangan holda ko‘plab bolalar nutqi korpuslari ishlab chiqildi. Quyidagi jadvalda ulardan eng muvaffaqiyatlilarini ko‘rishingiz mumkin.

	Korpus nomi	Muallif	Tillar	Ma‘ru-zachilar	Davo-miyligi	Yosh oralig‘i	Sana
1	CHILDES	Brayan MakVinni, Ketrin Snou	25 ta til			turlicha	1984
2	Demut Sesoto korpusi	Ketrin Demuth	Sesoto	4	98 soat	2-4	1992
3	TIDIGITS	R.Gary Leonard	Ingliz	326 (101 bola)		6-15	1993
4	CMU Kids Corpus	Maxine Eskinaze	Ingliz	76		6-11	1997
5	CU bolalar o‘qish va taklif qilingan nutq korpusi	Andreas Hagen	Ingliz	663		K – G5	2001
6	CU Story Corpus	Andreas Hagen	Ingliz	106	40 soat	G3–G5	2003
7	TABLO	Abe Kazemzadeh	Ingliz	256	40 soat	K-G4	2005

8	Shvetsiya NICE Korpusi	Linda Bell	Shvet	5580		8-15	2005
9	PF-STAR bolalar nutqi korpusi	Martin Rassel	Ingliz	158	14,5 soat	4-14	2006
10	Providence Corpus	Ketrin Demuth	Ingliz	6	363 soat	1-3	2006
11	CSLU Kids' Speech Corpus	Khaldoun Shobaki	Ingliz	1100		K-G10	2007
12	Lion korpusi	Ketrin Demuth	Fransuz	4	185 soat	1-3	2007
13	CHIEDE	Marta Garrote	Ispan	59	8 soat		2008
14	CASS_CHILD	Jun Gao	Mandarin	23		1-4	2012
15	CFSC	R.M.Paskual	Filippin	57	8 soat	6-11	2012
16	CALL-SLT	Manny Rayner	Nemis				2014
17	SingaKids – Mandarin	Nensi F.Chen	Mandarin	255	125 soat	7-12	2016
18	Boulder Learning-MyST Corpus	Cole va boshqalar	Ingliz	1371	393 soat	G3-G5	2019

Bolalar nutqi korpusi doirasida yaratilgan dastlabki loyiha – CHILDES korpusi edi. CHILDES (Child Language Data Exchange System) 1984-yilda Karnegi Mellon universitetining psixologiya bo‘limi doktorlari Brayan MakVinni va Ketrin Snoular tomonidan yaratilgan. Dastlab, ushbu dasturning maqsadi bolalar tilini o‘zlashtirish transkriptlari va yozuvlarini tadqiqotchilarga bepul, ommaviy axborot sifatida taqdim qilishdan iborat edi. Keyinchalik esa bu dastur korpus sifatida shakllantirildi. Uning eng dastlabki transkriptlari 1960-yillarga to‘g‘ri keladi. 1990-yillarning boshlarida CHILDESda topilgan ma‘lumotlardan osonlikcha foydalanish imkoniyati keng tarqalgani tufayli, undan foydalangan holda bolalar tilini o‘zlashtirish bo‘yicha yaratilgan tadqiqotlar soni sezilarli darajada oshdi.

CHILDES transkriptlarida turli tillar va yoshdagilar mavjud. Transkriptlarning aksariyati ssenariylashtirilmagan o‘zaro suhbatlardan iborat. Transkripsiyalar suhbat transkriptlarini ishlab chiqarish uchun standartlashtirilgan formatni ta‘minlovchi CHAT transkripsiya formatida kodlangan. Bu tizim har qanday til o‘rganuvchilar: bolalar, ikkinchi tilni o‘rganuvchilar va tiklanuvchi afaziklar bilan suhbatlarni transkripsiya qilish uchun ishlatilishi mumkin. Diskurs darajasidagi transkripsiyadan tashqari, CHAT tizimi fonologik va morfologik tahlil imkoniyatlariga ham ega. CHILDES asosan yosh bolalarning tilini tahlil qilish uchun ishlatiladi va kattalarning bolalar nutqiga munosabatiga qaratilgan.

Ayni paytda korpusda 30 ta tildagi 230 ta korpus mavjud bo‘lib, u nisbatan kattaroq TalkBank korpusining tarkibiy qismiga aylantirildi. Loyihada ma‘lumotlar bazasidan foydalangan 4000 nafar bolalar tili tadqiqotchilari va talabalar ro‘yxati mavjud hamda ushbu materiallardan foydalanish asosida chop etilgan 3000 dan ortiq maqolalar qayd etilgan. Loyiha, shuningdek, tildan foydalanishning turli darajalari, jumladan, leksika, sintaksis, morfologiya, fonologiya, nutq va hikoyani o‘rganish uchun foydali bo‘lgan kompyuter dasturlari to‘plamini yaratdi.

Bolalar nutqi korpusi borasida olib borilgan loyihalardan ahamiyatli sifatida Demut Sesoto korpusini ham keltirish mumkin. Ushbu korpus 1980-82-yillarda Afrikaning janubidagi Lesoto davlatida Ketrin Demuth tomonidan tuzilgan. Ushbu korpusni yaratishdan maqsad past xarajatli jarayon (ya‘ni, professional bo‘lmagan ko‘ngillilar va norasmiy yozib olish muhitlari yordamida yozilgan bir nechta ovozlardan foydalangan holda yuqori sifatli nutqni avtomatik aniqlash tizimini joriy etishni tekshirish edi. Ma‘lumotlar Moxotlong tumanidagi 550 kishidan iborat kichik Lesoto tog‘ qishlog‘ida to‘olangan. Natijada lug‘aviy fe‘llar, ½ million morfemalar va taxminan 13,250nta iborani o‘z ichiga olgan 98 soatlik nutq korpusi paydo bo‘ldi. Ushbu

materiallar qayta ishlangach CHILDES formatiga o'tkazildi va korpusning tarkibiy qismiga qo'shildi.

Bugungi kungacha yaratilgan va yaratilayotgan bolalar nutqi korpuslarining har biri nafaqat ta'lim, balki tibbiyot sohasidagi ko'plab muammolarga ham o'ziga xos yechimlar taqdim etmoqda. Bu jihatdan ularning har biri o'ziga xosdir. Biroq, ularning yaratilishiga asos bo'lgan CHILDES korpusi bugungi kunda ham yangi korpuslar uchun o'ziga xos qolip hamda boy ma'lumotlar bazasi vazifasini o'tamoqda desak, mubolag'a bo'lmaydi.

Foydalanilgan adabiyotlar:

1. Kodner, Jordan. 2019. Estimating child linguistic experience from historical corpora. *Glossa: a journal of general linguistics* 4(1): 122. 1–14. DOI: <https://doi.org/10.5334/gjgl.926>
2. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 10(8), 59-66.
3. Sharipbay, A., Shirinova, R., Abdurakhmonova, N., & Qodirova, Z. (2023). Ontological knowledge-based models representing medicine terminology. *Advances in Social Sciences and Economics*, 1(1), 4.
4. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). "Turkic Morpheme": From the Portal to the Linguistic Platform. In *World Conference Intelligent System for Industrial Automation* (pp. 181-189). Cham: Springer Nature Switzerland.
5. Dusmurod, N., Nilufar, N., Alimova, S., Abdurakhmonova, N., & Karabayeva, D. (2019). Social issues of rational nutrition. http://icame.uib.no/ij35/Hanhong_LI_and_Alex_C_Fang.pdf
6. <https://childes.talkbank.org/access/Other/Sesotho/Demuth.html>
7. <https://childes.talkbank.org/access/Eng-NA/0docs/MacWhinney1990.pdf>
8. <https://osf.io/zqb6c/wiki/home/>