Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

ЛІНГВІСТИЧНЕ АНОТУВАННЯ УСНОГО ДИТЯЧОГО МОВЛЕННЯ: ПРИНЦИПИ СТВОРЕННЯ СПЕЦІАЛІЗОВАНОГО КОРПУСУ

Кваліфікаційна робота
на здобуття ОС «бакалавр»
студентки IV курсу
галузі знань 03 «Гуманітарні науки»,
спеціальність 035 «Філологія»
спеціалізації 035.01 «Українська мова та література», ОПП «Українська мова і література та західноєвропейська мова»
Вероніки Михайлівни ГАЛІК

Науковий керівник:
асистент кафедри української мови та прикладної лінгвістики
Валентина РОБЕЙКО

«Допущено до захисту»
Протокол засідання кафедри
української мови та прикладної лінгвістики
від «05» червня 2026 року № 16

завідувач кафедри _____
к.філол.н., доц. Сергій РІЗНИК

КИЇВ – 2026

АНОТАЦІЯ

Бакалаврську роботу присвячено розробці та практичній реалізації процедури лінгвістичного анотування усного дитячого мовлення в програмному середовищі Praat для створення спеціалізованого корпусу, придатного до використання під час донавчання систем автоматичного розпізнавання мовлення (ASR). **Актуальність дослідження** зумовлена недостатньою точністю сучасних ASR-моделей при обробці дитячого мовлення та потребою у якісно розмічених даних для їх адаптації.

Об'єктом дослідження є спонтанне та модероване усне мовлення учнів 2-го класу, зафіксоване під час напівструктурованих інтерв'ю. **Предмет дослідження** становить процедура багаторівневого лінгвістичного анотування в Praat і мовленнєві явища, релевантні для навчання ASR-моделей. Методологічну основу роботи формують принципи корпусної лінгвістики, онтолінгвістики, експериментальної фонетики та соціолінгвістики. Для досягнення поставленої мети застосовано інструментальний аналіз у Praat, теоретичний, описово-аналітичний і статистичний методи.

Робота складається з трьох розділів. У першому розглянуто теоретичні засади створення мовних корпусів і особливості дитячого мовлення. У другому описано вибірку, розроблено правила транскрибування, чотирирівневу структуру TextGrid та систему маркерів для анотування. У третьому представлено практичну реалізацію анотування й аналіз отриманих результатів.

За підсумками дослідження сформовано корпус загальною тривалістю 76 хвилин 49 секунд, що містить 1 534 слова дитячих реплік і 203 ключові слова. Аналіз засвідчив середню частоту основного тону 271,9 Гц, що підтверджує акустичні відмінності дитячого мовлення від типових даних, на яких навчаються ASR-системи. Практичним результатом роботи стали розмічені .wav- та TextGrid-файли, готові до донавчання моделей, а також відтворювана система анотування, придатна для масштабування й використання у подальших дослідженнях.

Ключові слова: корпусна лінгвістика, усний корпус, дитяче мовлення, лінгвістичне анотування, Praat, TextGrid, автоматичне розпізнавання мовлення, ASR, донавчання, онтолінгвістика.

SUMMARY

This bachelor's thesis is devoted to the development and practical implementation of a procedure for the linguistic annotation of children's spoken language in the Praat software environment with the aim of creating a specialized corpus suitable for fine-tuning automatic speech recognition (ASR) systems. **The relevance of the study** is determined by the insufficient accuracy of modern ASR models in processing children's speech and the need for high-quality annotated data to facilitate their adaptation.

The object of the study is the spontaneous and moderated spoken language of second-grade pupils recorded during semi-structured interviews. **The subject of the study** is the procedure of multi-level linguistic annotation in Praat and speech phenomena relevant to the training of ASR models. The methodological framework is based on the principles of corpus linguistics, ontolinguistics, experimental phonetics, and sociolinguistics. To achieve the research objectives, instrumental analysis in Praat, as well as theoretical, descriptive-analytical, and statistical methods, were employed.

The thesis consists of three chapters. The first chapter examines the theoretical foundations of language corpus construction and the characteristics of children's speech. The second chapter describes the dataset, presents the developed transcription guidelines, the four-tier TextGrid structure, and the annotation marker system. The third chapter focuses on the practical implementation of the annotation procedure and the analysis of the obtained results.

As a result of the study, a corpus with a total duration of 76 minutes and 49 seconds was compiled, containing 1,534 words in children's utterances and 203 keywords. The analysis revealed a mean fundamental frequency of 271.9 Hz, confirming the acoustic differences between children's speech and the typical data used to train ASR systems. The practical outcomes of the thesis include annotated .wav and TextGrid files ready for model fine-tuning, as well as a reproducible annotation framework suitable for dataset expansion and application in future research.

Keywords: corpus linguistics, spoken corpus, children's speech, linguistic annotation, Praat, TextGrid, automatic speech recognition, ASR, fine-tuning, ontolinguistics.

ЗМІСТ

ВСТУП

Сучасне мовознавство є багатовекторним і поєднує власне вивчення структури, функціонування та розвитку мови із інноваційним інструментарієм і новітніми методами аналізу, розширюючи наукові горизонти для дослідників. Одним із найперспективніших розділів мовознавства є корпусна лінгвістика, завданням якої є створення, обробка та практичне використання мовних корпусів, зокрема спеціалізованих цифрових бібліотек або архівів.

На сучасному етапі розвитку комп'ютерної лінгвістики стрімко впроваджуються технології автоматичного розпізнавання мовлення (Automatic Speech Recognition, ASR), проте часто моделі не враховують вікові особливості мовців. Попри значні успіхи у розпізнаванні мовлення дорослих користувачів, моделі ASR досі демонструють суттєво нижчу точність під час обробки дитячого мовлення через його специфічні акустичні характеристики та лінгвістичні особливості. У межах прикладних проєктів, що працюють із великими масивами звукової інформації, це створює суттєве технологічне обмеження, адже низька якість розпізнавання критично уповільнює процес обробки вхідних аудіоданих, вимагаючи тривалого ручного корегування. Для вирішення цієї проблеми у межах прикладних лінгвістичних проєктів дедалі частіше створюють спеціалізовані (закриті) корпуси. Такі датасети призначені для вирішення локальних завдань — розпізнавання конкретного лексичного масиву та пришвидшення обробки первинних аудіоданих шляхом цільового донавчання моделей.

У зв'язку з цим **актуальність дослідження** зумовлена необхідністю створення цільового інструменту задля оптимізації внутрішніх робочих процесів та пришвидшення автоматизованої обробки великих обсягів аудіозаписів, виконаних за методикою ГО «Спільномова».

Хоча корпусна лінгвістика і відносно молода наука, навколо неї вже сформувалися **основні усталені позиції.** Питання корпусної лінгвістики, типізації та структурування текстових і усних масивів даних досліджували такі науковці, як Дж. Сінклер [Sinclair 1991], Т. Макінері та Е. Гарді [McEnery, Hardie

2011], а в українському мовознавстві — В. Широков [Широков 2005], Н. Дарчук +[Дарчук 2008] та інші. Проблематика онтолінгвістики та специфіки дитячого мовлення висвітлена у працях Дж. Брунера [Bruner 1983]., Д. Слобіна [Slobin 1985], а також вітчизняних дослідників психолінгвістичного напряму. Водночас перетин корпусного аналізу та інструментального анотування мовлення молодших школярів для навчання ASR-моделей в Україні залишаються недостатньо розробленими, що зумовлює наукову новизну роботи.

Мета дослідження — теоретично обґрунтувати та практично реалізувати процедуру створення й лінгвістичного анотування спеціального усного корпусу дитячого мовлення у програмному середовищі Praat для забезпечення подальшої оптимізації робочих процесів і донавчання моделей автоматичного розпізнавання мовлення (ASR).

Для досягнення поставленої мети визначено такі **завдання**:

1. Розглянути поняття лінгвістичного корпусу, охарактеризувати наявні типології усних корпусів;
2. Визначити специфічні особливості дитячого мовлення як об'єкта лінгвістичного опису та виявити чинники, що ускладнюють його автоматичне розпізнавання системами ASR;
3. Окреслити базові принципи розробки спеціалізованих (закритих) корпусів вузького призначення;
4. Охарактеризувати емпіричний матеріал дослідження та методику інтерв'ювання дітей, розроблену ГО «Спільномова»;
5. Систематизувати правила транскрибування та розробити робочі інструкції для текстової фіксації усного мовлення молодших школярів;
6. Описати технологію інструментального аналізу та виконати багаторівневу розмітку звукових сигналів у програмному середовищі Praat;
7. Здійснити процедуру сегментації та текстової транскрибації аудіофайлів, виявити й класифікувати особливості розмітки специфічних явищ дитячого мовлення;

8. Проаналізувати кількісні та якісні характеристики сформованого корпусу й оцінити потенціал його використання для донавчання розпізнавальних моделей.

Об’єктом дослідження виступає модероване та спонтанне усне мовлення дітей молодшого шкільного віку (учні 2-го класу, вік приблизно 7-8 років), зафіксоване в умовах інтерв'ю.

Предметом дослідження є процедура багаторівневого лінгвістичного анотування дитячого мовлення в програмі Praat та специфічні мовленнєві явища молодших школярів, релевантні для навчання систем ASR.

Методологічну основу дослідження становлять загальні принципи корпусної лінгвістики (репрезентативність, повнота, структурованість даних), базові положення онтолінгвістики та експериментальної фонетики, а також соціолінгвістичний підхід до аналізу мовленнєвої ситуації в освітньому просторі.

У роботі використано комплекс наукових **методів** для вирішення поставлених завдань, зокрема теоретичний аналіз наукових джерел і внутрішніх даних ГО «Спільномова», метод спостереження та польового запису у процесі інтерв’ювання, метод інструментально-лінгвістичного аналізу під час роботи у програмі Praat, описово-аналітичний метод для систематизації виявлених мовленнєвих явищ, а також елементи статистичного методу для підрахунку обсягу отриманих текстових і часових сегментів корпусу.

Наукова новизна роботи полягає в тому, що вперше для потреб ГО «Спільномова» теоретично обґрунтовано та практично реалізовано процедуру багаторівневого лінгвістичного анотування спеціального усного корпусу дитячого мовлення в середовищі Praat; систематизовано специфічні явища мовлення молодших школярів, актуальні для навчання ASR-моделей, і розроблено інструкції для їх транскрибування та розмітки.

Практичне значення отриманих результатів полягає у безпосередньому формуванні бази лінгвістичних даних – текстових та акустичних матеріалів (розмічених аудіофайлів у форматі .wav та відповідних шарів анотації TextGrid у програмі Praat), які готові до інтеграції в процеси донавчання моделей

автоматичного розпізнавання українського мовлення. Розроблені інструкції та принципи лінгвістичного коментування можуть бути використані під час розширення цього чи проєктування аналогічних усних корпусів.

Структура роботи. Кваліфікаційна робота складається зі вступу, трьох розділів, висновків, списку використаних джерел і додатків. Загальний обсяг роботи становить *45 сторінок основного тексту; список використаних джерел містить 46* позицій.

РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ СТВОРЕННЯ УСНИХ ЛІНГВІСТИЧНИХ КОРПУСІВ

Проведення будь-якого прикладного лінгвістичного дослідження вимагає ґрунтовної теоретичної основи. Без розуміння того, що таке корпус, чим усне мовлення відрізняється від письмового і чому дитячий голос є складним об'єктом для автоматичного розпізнавання, неможливо обґрунтувати ані методику анотування, ані архітектуру розмітки, ані вибір інструментарію. Тому перший розділ кваліфікаційної роботи присвячено теоретико-методологічним засадам, на яких ґрунтується практична частина дослідження.

Розділ структуровано за логікою поступального звуження предмета: від загального поняття лінгвістичного корпусу та типології усних корпусів (підрозділ 1.1) — до специфічних особливостей дитячого мовлення як об'єкта лінгвістичного опису й автоматичного розпізнавання (підрозділ 1.2), і далі — до конкретних принципів розробки спеціалізованих закритих корпусів для оптимізації ASR-систем (підрозділ 1.3).

1.1. Поняття лінгвістичного корпусу та типологія усних корпусів

Сучасне мовознавство розвивається надзвичайно швидко. Однією з найперспективніших його галузей є корпусна лінгвістика, яка займається формуванням і опрацюванням корпусів — свого ключового поняття [Карпіловська 2014]. Незважаючи на широке вживання терміна, дослідники пропонують різні його тлумачення. В. А. Широков у монографії «Корпусна лінгвістика» визначає корпус як репрезентативну машинозчитувану колекцію текстів природної мови, укладену відповідно до певних лінгвістичних критеріїв та забезпечену необхідною анотацією [Широков 2005, с. 21]. Схожу позицію займає Н. П. Дарчук, акцентуючи на тому, що в комп'ютерній лінгвістиці корпус є повноцінним інструментом автоматичного опрацювання мовного матеріалу, а не просто зібранням текстів [Дарчук 2008, с. 36].

Одним із перших у зарубіжному мовознавстві фундаментальних визначень корпусу вважається формулювання Дж. Сінклера, за яким корпус — це зібрання текстів природного походження, відібраних і впорядкованих відповідно до

заздалегідь визначених мовних критеріїв [Sinclair 1991, с. 14]. Пізніше сам Дж. Сінклер уточнив, що принципова відмінність корпусу від архіву полягає у наявності структурованих метаданих, їхнього опису та лінгвістичної розмітки, яка робить зібрання придатним для систематичного аналізу [Sinclair 2005, с. 9]. Т. Макінері та Е. Гарді розглядають корпус як ресурс, де репрезентативність, машинозчитуваність і цілеспрямованість відбору є взаємозумовленими, а не незалежними ознаками [McEnery, Hardie 2011, с. 2-14]. Г. Кеннеді наголошує, що коректно укладений корпус відображає реальне функціонування мови в різноманітних ситуаціях спілкування, а не є штучною вибіркою [Kennedy 1998, с. 62].

В українській традиції корпусної лінгвістики важливими є напрацювання О. Демської-Кульчицької, яка в своїх роботах виокремлює такі обов'язкові ознаки корпусу, як:

— автентичність мовного матеріалу;

— репрезентативність вибірки;

— наявність лінгвістичної розмітки;

— машинночитність даних.

Дослідниця також підкреслює, що масив нерозмічених текстів або аудіозаписів, яким бракує хоча б однієї з цих властивостей, залишається лише архівом і не є повноцінним корпусом [Демська-Кульчицька 2003, с. 41].

В. В. Жуковська підсумовує вищеназвані позиції й пропонує визначати корпус як «машиночитане, збалансоване, репрезентативне зібрання особливо розмічених (анотованих) текстів, відібраних згідно фіксованих параметрів для досягнення визначеної лінгвістичної мети та досліджуваних нелінійно за принципом гіпертексту» [Жуковська 2013, с. 58]. На важливості опрацювання й розмітки корпусу акцентує більшість дослідників, зокрема і Т. В. Бобкова наголошує на тому, що аналітичний потенціал корпусу безпосередньо визначається якістю його анотаційного оформлення [Бобкова 2014, с. 16].

Принципова відмінність усних корпусів від текстових полягає в природі первинного носія даних. Письмовий текст є дискретним об'єктом, адже в ньому

уже графічно позначені межі слів, речень і абзаців. Натомість усне мовлення — це безперервний звуковий потік, позбавлений природних пробілів і розділових знаків [Thompson 2005, с. 73]. Усне мовлення розгортається в часі й несе у собі просодичні характеристики (тон, темп, інтенсивність), що суттєво впливають на змістове навантаження висловлювань і не можуть бути відтворені у звичайному буквеному записі [McEnery, Hardie 2011, с. 2-14]. Таким чином сегментація потоку мовлення на дискретні одиниці (слова, фрази, репліки) є як суто технічним, так і аналітичним актом, адже вимагає одночасного застосування слухового досвіду, знань із фонетики й просодики, а також використання наперед попередньо визначеної розмітки. Не дарма В. В. Жуковська вказує, що саме процедура транскрипції з'єднує мовленнєвий запис і лінгвістичну анотацію, перетворюючи аудіоархів на повноцінний корпус усного мовлення [Жуковська 2013, с. 76].

Загальноприйнятої класифікації усних і текстових корпусів наразі немає, тож будь-яка їхня типізація є розгалуженою і будується за кількома незалежними параметрами. Питання класифікації текстових і мовленнєвих корпусів О. Демська-Кульчицька розглядає окремо, вказуючи на різноманіття підстав для типологізації залежно від дослідницьких цілей. Так, наприклад, окрім розрізнення на усні й письмові, виділяють повнотекстові, фрагментні, дослідницькі, ілюстративні, фундаментальні, динамічні, статичні, діахронні, синхронні, інтерпретаційні, загальномовні, спеціалізовані, паралельні, порівняльні, неанотовані і анотовані корпуси [Демська-Кульчицька 2004, с. 154-157].

Якщо говорити винятково про усні корпуси, то їх варто розрізняти не лише за загальними критеріями, а й за характером мовленнєвого матеріалу (читане, підготовлене та спонтанне мовлення), за формою комунікації та кількістю учасників мовленнєвої взаємодії (монологічні, діалогічні та полілогічні), за віком мовців (дитячі, підліткові, дорослі, змішані), за статтю, регіоном проживання, соціальними чи професійними характеристиками тощо. З огляду на використання персональних даних і голосу окремих мовців, а також з етичних,

правових чи організаційних міркувань, для усних корпусів особливо актуальним є розрізнення на відкритого й закритого типу доступу.

До найвідоміших у світовій науці прикладів усних корпусів належить CHILDES (Child Language Data Exchange System) — міжнародна база даних транскриптів дитячого мовлення, що на сучасному етапі активно розвивається в межах глобальної мультимедійної системи TalkBank. На сьогодні цей проєкт охоплює десятки мов світу й містить унікальну колекцію моно- та білінгвальних записів різного хронометражу, а завдяки суворому дотриманню стандартів маркування слугує еталонним і найбільш авторитетним ресурсом в онтолінгвістичних та психолінгвістичних дослідженнях [MacWhinney 2000, с. 5-13].

Важливим кроком у розвитку корпусів спонтанного, зокрема дитячого, мовлення став репозиторій HomeBank, який є суміжним та інтегрованим із CHILDES науковим проєктом. Головна специфіка цього сховища полягає у збиранні та довготривалому зберіганні цілодобових аудіозаписів природного мовлення. Усі матеріали для бази даних записуються безпосередньо у звичному сімейному оточенні дітей без штучного втручання дослідників і фіксують унікальні акустичні параметри щоденного мовлення [VanDam 2016, с. 133-137].

Сучасний підхід до представлення усних даних реалізовано на європейській платформі Spokes [Spokes 2025], яка діє як вебсервіс і забезпечує користувачам швидкий пошук, візуалізацію та перегляд корпусів, зокрема підкорпусу польського спонтанного мовлення, на базі PELCRA та CLARIN-PL [Pęzik 2015, с. 99-101].

1.2. Особливості дитячого мовлення як об'єкта лінгвістичного опису й автоматичного розпізнавання (ASR)

Дитяче мовлення утворює специфічну мовленнєву систему, що є водночас предметом онтолінгвістики, соціолінгвістики та комп'ютерної фонетики. Молодший школяр 7–8 років (учень 2 класу) перебуває на важливому перехідному етапі мовленнєвого розвитку: він уже оволодів базовими фонологічними та граматичними структурами рідної мови, однак цей процес ще

не завершений. За спостереженнями Дж. Брунера, саме шкільний вік є критичним для формування комунікативних та наративних стратегій, тому мовлення дітей у цей період відзначається одночасно і активним розширенням мовних засобів, і значною нестабільністю їхнього застосування [Bruner 1983, с. 17-19]. М. Томаселло показав, що раннє синтаксичне становлення спирається не на абстрактні граматичні правила, а на засвоєні конкретні конструкції (item-based constructions), що дозволяє припустити численні незавершені фрази, раптові зміни теми й самовиправлення в мовленні молодших школярів [Tomasello 2000, с. 156-163].

Словниковий запас дітей 7–8 років активно зростає, однак між пасивним і активним лексиконом зберігається суттєвий розрив. Для оцінки словникового запасу в міжнародній практиці застосовуються стандартизовані інструменти, наприклад, PPVT-5 (Peabody Picture Vocabulary Test, п'яте видання) вимірює обсяг рецептивного (пасивного) лексикону шляхом пред'явлення ілюстрацій [Dunn 2018, Dunn 2019, с. 1-5], тоді як EVT-3 (Expressive Vocabulary Test, третє видання) спрямований на оцінку продуктивного (активного) словника [Williams 2018, с. 1, Williams 2018, с. 1-2]. Для оцінки наративних здібностей, зокрема у білінгвів, розроблено MAIN (Multilingual Assessment Instrument for Narratives), що дозволяє порівнювати мовленнєвий розвиток дітей у різних мовних середовищах [MAIN 2026, Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, Walters 2015, с. 2-24]. Ці методологічні рамки суттєво вплинули на дизайн методики збирання матеріалу для описуваного корпусу: завдання на переказ, опис картинок і складання речень безпосередньо корелюють з принципами еліситації мовлення, прийнятими у зазначених інструментах.

Специфіка мовного середовища Києва додає до лінгвістичного портрету другокласника соціолінгвістичний вимір. В умовах тривалої українсько-російської двомовності частина дітей є природними білінгвами або мовцями із суржиковими вкрапленнями. Для таких мовців характерне явище інтерференції на рівні фонетики, лексики та граматики, а також кодових переключень у межах одного висловлювання. Ф. Бацевич Ф. у «Основах комунікативної лінгвістики»

розглядає змішане мовлення як закономірний наслідок функціонування в контакті двох близькоспоріднених мов [Бацевич 2004, с. 251-267]. Система розмітки корпусу враховує цю особливість, передбачаючи спеціальні маркери для позначення повністю російськомовних сегментів (*р*) і суржикових включень (*сур*).

Акустико-фонетичні особливості дитячого голосу є безпосередньою причиною неефективності стандартних ASR-систем при обробці дитячого мовлення. Анатомічно коротші голосові зв'язки і менший об'єм голосового тракту у дітей порівняно з дорослими зумовлюють значно вищу частоту основного тону (F0). Тобто, якщо для дорослих чоловіків середня F0 становить приблизно 120 Гц, а для жінок — 210 Гц, то у дітей до початку пубертату вона типово перебуває в діапазоні 220–300 Гц і вище [Average Speaking Frequencies 2026]. Формантна структура голосних у дитячому мовленні також суттєво відрізняється: значення F1 і F2 розташовані вище, а міжмовленнєва варіативність значно більша навіть у межах одного запису [Gerosa1, Giuliani, Narayanan, Potamianos 2009, с. 2-3]. Все це разом формує акустичний профіль, принципово відмінний від того, на якому тренуються промислові системи розпізнавання мовлення.

Проблему акустичного дисонансу (acoustic mismatch) між дитячим і дорослим мовленням описано у фундаментальному огляді М. Ґерози, Д. Джуліані, Ш. Нараянана та О. Потаміаноса [Gerosa1, Giuliani, Narayanan, Potamianos 2009, с. 1-7]. Автори констатують, що сучасні ASR-моделі тренуються на масивах записів дорослих мовців (наприклад, на записах мовлення дикторів, аудіокнигах, подкастах), і виявляються структурно несумісними з акустичними характеристиками дитячого голосу. Систематичний огляд В. Бгарваджа та ін. підтверджує, що показник помилок Word Error Rate (WER) для дитячого мовлення залишається стабільно вищим порівняно з дорослим навіть у найпотужніших сучасних архітектурах [Bhardwaj, Othman, Kukreja, Belkhier, Bajaj, Srikanth Goud, Ur Rehman, Shafiq, Hamam 2022, с. 18].

Паралінгвістичні явища, а саме сміх, дихання, шепіт, тривалі паузи роздумів і хезитації, є додатковим чинником збоїв алгоритмів, адже система або

хибно декодує їх як лексичні одиниці, або ігнорує суміжні слова через порушення очікуваної моделі мовленнєвого потоку [Gerosa1, Giuliani, Narayanan, Potamianos 2009, с. 2-6, Bhardwaj, Othman, Kukreja, Belkhier, Bajaj, Srikanth Goud, Ur Rehman, Shafiq, Hamam 2022, с. 2-12]. Дослідження О. С. Іщенко про вплив частин мови на паузування в усному мовленні дорослих [Іщенко 2020, с. 45-55] дозволяє висунути припущення, що і в дитячих паузах є певна системність, і це явище пов'язане з когнітивним навантаженням при виборі лексичних одиниць.

1.3. Принципи розробки спеціалізованих (закритих) корпусів для оптимізації систем автоматичного розпізнавання мовлення (ASR)

Пошук рішення для проблеми низької точності ASR на дитячому мовленні наштовхує на концепцію спеціалізованого корпусу. Великі загальнонаціональні мовленнєві бази містять широкий спектр голосів, тематик і реєстрів, однак вони не можуть відображати акустику та лексику конкретної вікової групи в конкретному регіональному та соціолінгвістичному контексті [McEnery, Hardie 2011, с. 2-14, Демська-Кульчицька 2003, с. 41-42]. Натомість так звані ad-hoc корпуси (масиви даних, цілеспрямовано зібрані для розв'язання однієї конкретної задачі) дозволяють сконцентрувати анотаційний ресурс саме там, де він потрібен. Дж. Сінклер підкреслює, що будь-яке рішення про склад корпусу визначається наперед сформульованою дослідницькою або прикладною метою, і тому відбір матеріалу завжди повинен бути теоретично обґрунтованим [Sinclair 2005, с. 15-19]. Г. Кеннеді, розглядаючи різноманіття корпусних проєктів, вказує, що небагато правильно проанотованих прикладів часто дають системі більше, ніж тисячі нерозмічених записів [Kennedy 1998, с. 60-63].

Прикладним контекстом цього дослідження є діяльність ГО «Спільномова» — громадської організації, що займається підтримкою україномовного середовища через освітні та технологічні проєкти [Спільномова 2023]. За даними дописів засновника організації Андрія Ковальова, починаючи з кінця 2023 року, ГО «Спільномова» послідовно здійснює записи мовлення дітей в межах дослідження їхнього рівня володіння українською мовою [Допис Andriy Kovalyov (Facebook) 2023, Допис Andriy Kovalyov (Facebook) 2024, Допис Andriy

Kovalyov (Facebook) 2024]. Кількість охочих пройти дослідження зростає, проте ресурси ГО обмежені, тому виникла ідея спростити й хоча б частково автоматизувати процес розпізнавання мовлення на зібраних аудіофайлах. Ця практична потреба напряму визначає вимоги до корпусу, адже він не має бути широким, натомість повинен точно опрацьовувати обмежену лексику в обмеженій віковій категорії, тобто може бути малим за обсягом, але досконалим щодо якості анотацій і відповідати поняттю «золотого датасету» (gold standard dataset).

Технологія донавчання (fine-tuning) є лінгвістично релевантним процесом, навіть якщо його математична сторона залишається поза межами цього дослідження. Базові мультимовні ASR-моделі, такі як OpenAI Whisper та wav2vec 2.0, вже можуть загально опрацьовувати українську мову, адже вони навчені на величезних об'ємах мовленнєвих даних і здатні розпізнавати вимову дорослих мовців [Radford, Kim, Xu, Brockman 2022, с. 1-4, Baevski, Zhou, Mohamed, Auli 2020, с. 1-8]. Однак для того, щоб модель адаптувалася до акустики 7–8-річних дітей конкретного середовища та до специфічної лексики дослідницького інтерв'ю, потрібне адаптаційне тренувальне доповнення. Воно формується на основі детально проанотованого корпусу, де кожен відрізок аудіосигналу зіставлений з текстовим відповідником. Вищеописаний систематичний огляд попередніх напрацювань і розробок підтверджує, що навіть невеликий специфічний датасет здатен значно знизити WER при донавчанні порівняно з базовою моделлю.

Можемо стверджувати, що архітектура багаторівневої розмітки визначається функціональними вимогами ASR. Система розпізнавання мовлення потребує рівня, де буде вказано нормативний текст, тобто те, що вимовив мовець (у цьому випадку, дитина). Поруч із ним необхідно вести систему розмітки, що фіксує паузи, шуми, хезитації та нестандартні форми, адже без цього рівень, який містить лише «чисті» слова, не дає системі уявлення про реальну акустичну картину. З урахуванням того, що розроблюваний корпус повинен виконувати специфічні задачі, а саме не лише розпізнавати мовлення, а й шукати в ньому

слова з обмеженого переліку, пріоритетним є той рівень, що розрахований на виокремлення ключових слів. Такий розподіл відповідає логіці, прийнятій у сучасній практиці підготовки мовленнєвих датасетів [Lennes 2009, с. 4-5].

З урахуванням того, що об’єктом корпусу є мовлення неповнолітніх осіб, під час розробки корпусу виникають правові обмеження. Голос людини є біометричною інформацією і підпадає під захист відповідно до Закону України «Про захист персональних даних» [Закон України «Про захист персональних даних» 2010]. Додаткові вимоги щодо роботи з даними неповнолітніх регламентовані окремими нормами [Захист персональних даних неповнолітніх 2025]. Запис дитячого мовлення вимагає попередньої інформованої письмової згоди батьків або законних представників, а зберігання і подальше використання матеріалів мають відповідати принципу мінімізації даних і цільового обмеження. Аналогічні принципи закритого та некомерційного зберігання дитячих аудіоданих описані у роботі М. ВанДама щодо репозиторію HomeBank: автори наголошують, що проблема розподілу та захисту даних є одним із ключових методологічних викликів при роботі з корпусами записів дитячого мовлення [VanDam 2016, с. 133-137]. Таким чином, закритий статус описуваного в цій роботі корпусу є усвідомленим правовим і етичним рішенням.

Висновки до першого розділу

Перший розділ присвячено теоретико-методологічним засадам створення усних лінгвістичних корпусів. Аналіз наукових праць вітчизняних і зарубіжних дослідників (В. А. Широкова, Н. П. Дарчук, О. Демської-Кульчицької, В. В. Жуковської, Дж. Сінклера, Т. Макінері та Е. Гарді, Г. Кеннеді) дозволив встановити, що лінгвістичний корпус є системно зорганізованим, репрезентативним і обов'язково анотованим масивом мовних даних, придатних для автоматизованого опрацювання. Обов'язковими ознаками повноцінного корпусу визнано автентичність мовного матеріалу, репрезентативність вибірки, наявність лінгвістичної розмітки та машиночитність.

Встановлено, що принципова відмінність усних корпусів від письмових зумовлена безперервністю звукового потоку, який позбавлений природних

пробілів і розділових знаків та вимагає складної процедури часової сегментації на дискретні одиниці з обов'язковим фіксуванням просодичних та акустичних характеристик. Транскрипція є необхідною сполучною ланкою між первинним аудіоархівом і повноцінною лінгвістичною анотацією. Огляд типології засвідчив широке варіювання підходів до класифікації корпусів: за характером мовленнєвого матеріалу, формою комунікації, кількістю учасників, віком мовців та типом доступу. Для цього дослідження ключовою є диференціація спеціалізованих (закритих, ad-hoc) корпусів, орієнтованих на розв'язання конкретних прикладних завдань. Міжнародний досвід систем CHILDES, HomeBank і Spokes підтвердив, що аналітичний і прикладний потенціал корпусу безпосередньо визначається якістю лінгвістичної розмітки.

Охарактеризовано специфіку дитячого мовлення як складного об'єкта лінгвістичного опису та автоматичного розпізнавання. Встановлено, що молодший школяр 7–8 років перебуває на перехідному етапі мовленнєвого розвитку: базові фонологічні та граматичні структури вже засвоєні, однак нестабільність їхнього застосування зберігається. Для цієї вікової групи характерними є суттєвий розрив між активним і пасивним лексиконом, вживання item-based конструкцій (за М. Томаселло), незавершені синтаксичні структури, самовиправлення та часті хезитації. Окремим релевантним чинником у контексті мовлення київських другокласників є явища інтерференції та кодових переключень між українською і російською мовами, характерні для білінгвальних та суржикомовних мовців.

Акустико-фонетичний аналіз засвідчив, що дитячий голос має принципово відмінний від дорослого акустичний профіль: середня частота основного тону у дітей перевищує 220–300 Гц порівняно з 120–210 Гц у дорослих, а формантна структура голосних зазнає характерного зміщення вгору. Ці особливості формують проблему «акустичного дисонансу» (acoustic mismatch), яка є ключовою причиною стабільно вищого рівня помилок Word Error Rate (WER) при обробці дитячого мовлення сучасними ASR-системами, навченими переважно на мовленні дорослих мовців.

Обґрунтовано принципи розробки спеціалізованих (закритих) корпусів вузького призначення. Встановлено, що цільовий «золотий датасет» (gold standard dataset), зосереджений на конкретній лексиці, обмеженій віковій та регіональній групі, є ефективнішим інструментом для адаптаційного донавчання (fine-tuning) базових ASR-моделей (зокрема Whisper та wav2vec 2.0), ніж великі загальнонаціональні мовленнєві бази. Архітектуру багаторівневої розмітки визначено як функціонально зумовлену вимогами ASR: система потребує паралельних рівнів нормативного тексту, паралінгвістичних явищ та цільових ключових слів. Закритий статус корпусу обґрунтовано як усвідомлене правове й етичне рішення, що відповідає вимогам Закону України «Про захист персональних даних» стосовно обробки біометричних даних неповнолітніх осіб.

РОЗДІЛ 2. МЕТОДИКА ТА ІНСТРУМЕНТАЛЬНІ ЗАСОБИ ЛІНГВІСТИЧНОГО АНОТУВАННЯ ДИТЯЧОГО МОВЛЕННЯ

Теоретичні засади, викладені в першому розділі, потребують конкретизації через опис матеріалу, методів і технічних засобів роботи. Другий розділ виконує цю функцію, відповідаючи на питання, з чим саме працює дослідник, за якими правилами повинен бути зібраний і розмічений мовленнєвий матеріал та в якому програмному середовищі реалізується анотування.

Структура розділу включає опис трьох послідовних аспектів методичної роботи. У підрозділі 2.1 схарактеризовано емпіричну базу дослідження — умови збирання аудіозаписів та критерії відбору матеріалу для фінальної вибірки. Підрозділ 2.2 присвячено розробці системи транскрибування, обґрунтовано базові принципи фіксації мовлення і детально описано систему анотаційних маркерів. Підрозділ 2.3 описує технологію роботи в програмі Praat, а саме архітектуру файлу TextGrid і конкретні аналітичні функції, що були задіяні для анотування дитячого мовлення.

2.1. Характеристика емпіричного матеріалу та умов його збирання

Емпіричну базу дослідження складають аудіозаписи, зібрані з квітня по травень 2024 року на базі Ліцею № 107 «Введенський» міста Києва в межах дослідницького проєкту ГО «Спільномова» [Спільномова 2023]. Записи у форматі індивідуального напівструктурованого інтерв'ю здійснювали заздалегідь проінструктовані модератори дитячого мовлення — студенти 2 року навчання першого (бакалаврського) ступеня вищої освіти. Загальний обсяг матеріалу, зібраного в той період у Ліцеї № 107 «Введенський» міста Києва, становить 67 аудіозаписів тривалістю кожного від 20 до 30 хвилин, із яких 37 — зразки мовлення дітей 1-го класу, а решта 30 належать учням 2-го класу. Записи здійснювалися на диктофон особистих мобільних телефонів модераторів, зберігалися у форматі .m4a і в кінці кожного дня завантажувалися у спільне хмарне сховище «Google Drive».

Під час формування вибірки для подальшого лінгвістичного анотування застосовувалися критерії відбору, що забезпечили однорідність умов і репрезентативність матеріалу. Відхилялися записи, у яких:

а) рівень фонового шуму (галас у коридорі під час перерви, гучний шкільний дзвінок, скрип меблів, сторонні голоси) суттєво ускладнював розбірливість мовлення;

б) методика не була застосована в повному обсязі через брак часу або втрату концентрації дитиною;

в) досвід модератора був недостатнім, що виявлялось у відступах від прийнятого протоколу.

Крім того, відбиралися виключно записи, зроблені безпосередньо у м. Київ (а не в Київській чи інших областях), що забезпечувало мовну однорідність соціолінгвістичного контексту. Подібна вимогливість до якості матеріалу узгоджується з вимогами до gold standard датасету, де точність кожного зразка важливіша за загальний обсяг [Simmering 2024].

До фінальної вибірки увійшли три записи учнів 2Б класу: двох дівчаток (Віка К. та Кіра Н.) і одного хлопчика (Юхим К.). Відібрані мовці утворюють контрастивну тріаду за мовним профілем: Віка К. є переважно українськомовною дитиною з використанням англійської у медіасередовищі, Кіра Н. перебуває в середовищі активного українсько-російського білінгвізму, а Юхим К. формується у фактично повністю російськомовному середовищі й повсякденні. Таке різноманіття відображає реальний мовний ландшафт київських шкіл і є свідомим рішенням для максимальної репрезентативності при подальшому донавчанні систем ASR . Усі троє дітей мали заповнені батьками анкети, що містять відомості про мову спілкування дитини вдома, у школі, з однолітками, а також про мову медіаспоживання (задля зручності дані з опитування батьків скорочено відображено у таблиці (табл. 2.1.).

Таблиця 2.1

Ім'я та перша літера прізвища дитини (наприклад, Матвій К.)	Якою є домінантна мова дитини (мова якою дитина говорить більшу частину часу, продукує спонтанні висловлювання, якою дитина переважно думає)?	Якою мовою розмовляє мама/тато з дитиною?	Якою мовою розмовляє третій значимий дорослий з дитиною?	Якою мовою розмовляють брат та/або сестра з дитиною?	Якою мовою дитина споживає контент на ютьюбі (мультфільми, ролики, тощо)?	Якою мовою дитина грає в комп'ютерні ігри?	Яку мову дитина переважно чує від інших дітей при взаємодії з ними на вулиці, дитячих майданчиках?
Юхим К.	Російська	Переважно російською + українська (75% або більше рос, 25% або менше українською)	Лише російською (практично 100% часу)	Немає	Переважно російською + українська (75% або більше часу - рос, 25% або менше часу - укр)	Лише російською (практично 100% часу)	Українську та російську однаковою мірою (50% часу - укр, 50% часу - рос)
Вікторія К	Українська	Лише українською (практично 100% часу)	Лише українською (практично 100% часу)	Лише українською (практично 100% часу)	Українською та англійською однаковою мірою (50% часу - укр, 50% часу - англ)	Українською та англійською однаковою мірою (50% часу - укр, 50% часу - англ)	Переважно українську + російську (75% або більше часу - укр, 25% або менше часу - рос)
Кіра Н.	Українська	Переважно українською + російська (75% або більше часу - укр, 25% або менше часу - рос)	Переважно українською + російська (75% або більше часу - укр, 25% або менше часу - рос)	Переважно українською + російська (75% або більше часу - укр, 25% або менше часу - рос)	Українською та російською однаковою мірою (50% часу - укр, 50% часу - рос)	Українською та російською однаковою мірою (50% часу - укр, 50% часу - рос)	Українську та російську однаковою мірою (50% часу - укр, 50% часу - рос)

Умови проведення запису були стандартизовані: модератор запрошував дітей по одному до окремого класного приміщення, де вони сиділи за партою один навпроти одного. Методика проведення інтерв'ю була розроблена колективом ГО «Спільномова» з урахуванням міжнародного досвіду оцінки мовленнєвого розвитку, зокрема за умов білінгвізму. В основу завдань, які були запропоновані дітям до виконання, лягли загально визнані напрацювання таких сертифікованих і перевірених на практиці в інших країнах тестів, як PPVT-5 (Peabody Picture Vocabulary Test) [Dunn 2018, Eigsti 2017 c. 1-4, Dunn 2019, c. 1-5], EVT-3 (expressive vocabulary test) [37, 38] [Williams 2018, c. 1, Williams 2018, c. 1-2], MAIN (multilingual assessment instrument for narratives), [MAIN 2026, Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, Walters 2015, c. 2-24].

Інтерв'ювання складалося з п'яти послідовних блоків:

1. пояснення значення на контрольних слів із тексту (перевірка ізольованих лексем);

2. переказ прочитаного модератором тексту по абзацах (перевірка розуміння прочитаного через контрольні події);
3. складання довільних речень за картками з ілюстраціями предметів-ключових слів (розмір карток 10×7 см);
4. вільний опис одного ілюстрованого розвороту А3 з книги «Зоопарк! Віммельбух» Каролін Гертлер [46] [Гертлер 2016, с. 3];
5. перевірка ключових слів, не названих дитиною самостійно, або названих російською.

Такий формат спілкування з дитиною поєднує різні режими використання мови — від цілеспрямованого називання до розгорнутого монологічного опису — і аналогічний за структурою до завдань, що застосовуються в PPVT та EVT [Dunn 2018, Eigsti 2017 с. 1-4, Williams 2018, с. 1, Williams 2018, с. 1-2], а також у методиці MAIN для дослідження наративних здібностей білінгвів [MAIN 2026, Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, Walters 2015, с. 2-24]. Важливо, що ані батьки, ані вчителі, ані модератори не могли повідомляти опитуваним суть дослідження чи давати такі настанови щодо виконання завдань, які могли б розкрити її («Зараз перевірятимуть твою українську», «Як це сказати українською?»).

За цих умов основним завданням модератора є підтримка еліситації, яка досягається завдяки фасилітації. Останню забезпечують такі аспекти:

1. етичне маркування аудіозаписів;
2. дотримання послідовності завдань;
3. використання запропонованих матеріалів;
4. розподіл часу на виконання завдань;
5. контроль сказаного дитиною;
6. контроль сказаного модератором;
7. підтримка морально-психологічного стану дитини;
8. індивідуальний підхід та адаптація завдань.

До зовнішніх чинників, що вплинули на якість деяких записів, належать:

1. фоновий шум з коридору;

2. одиночні або тривалі звуки шкільного дзвінка;
3. скрип меблів;
4. шум від використання дитиною сторонніх предметів;
5. несанкціонований вхід до класу;
6. переривання через потреби дитини;
7. переривання через оголошення повітряної тривоги.

Частина цих факторів є об'єктивно нездоланними за умов польового запису у реальному шкільному середовищі та фіксується у системі розмітки спеціальними маркерами (*шт*, *скр*, *крикд*, *шар*). Аналогічні виклики польового збирання дитячих аудіоданих описуються у контексті репозиторію HomeBank, де наголошується, що контроль акустичних умов у натуральних умовах запису завжди є частковим [VanDam 2016, с. 133-137].

2.2. Правила транскрибування та інструкції для текстової фіксації усного мовлення молодших школярів

Основним завданням системи транскрибування є максимально точне відтворення реально сказаного при одночасному дотриманні прийнятих орфографічних норм. Під час розробки інструкцій для текстової фіксації усного мовлення молодших школярів було враховано фактичні показники мовлення опитаних дітей, а також імовірні виклики, які могли б виникнути на наступних етапах роботи із проанотованим матеріалом.

За базовий принцип прийнято, що мова запису є українська, а всі слова записуються відповідно до чинних норм сучасної української літературної мови [Правопис 2019]. Відхилення від норми фіксуються не довільною фонетичною транскрипцією, яка б лише ускладнила машинне опрацювання і зробила б корпус несумісним зі стандартними ASR-конвеєрами, а за допомогою спеціальної системи маркерів, що виділяються з обох боків астериском (*) і ставляться перед тим словом, якого стосуються. Такий підхід збігається із загальними принципами анотування усного мовлення, описаними в корпусній лінгвістиці [Демська-Кульчицька 2003, с. 41-42, Lennes 2009, с. 4-5].

Згідно з технічним завданням, для описуваного корпусу обрано архітектуру з чотирма паралельними інтервальними рівнями (interval tiers), що дозволяє одночасно відображати і зіставляти мовлення обох учасників взаємодії у єдиній часовій осі, забезпечуючи коректну фіксацію епізодів одночасного мовлення та перебивань. Виокремлення саме інтервальних рівнів замість точкових (point tiers) зумовлене лінгвоакустичною природою усного мовлення, яке розгортається в часі й потребує точної фіксації часових меж (початку та кінця) кожної звукової одиниці. Принцип паралельних рівнів у розмітці мовленнєвих корпусів обстоюється у роботах М. Леннеса, зокрема при описі методології аналізу спонтанного і читаного мовлення у фінській фонетиці [Lennes 2009, с. 4-5] та при розробці пакету SpeCT 2.0 для Praat [Lennes 2019, с. 2-4].

Сформована архітектура розмітки передбачає такі рівні (tiers):

Tier 1 — мовлення модератора;

Tier 2 — мовлення дитини;

Tier 3 — ключові слова, названі модератором;

Tier 4 — ключові слова, названі дитиною.

Усі паузи, шуми, просодичні та паралінгвістичні позначки фіксуються на тому рівні, де вони реально виникають, і синтаксично інтегруються в текст відповідного сегменту. Наприклад, якщо під час відповіді дитини заскрипів стілець, дитина видихнула і сказала «Добре» — запис матиме вигляд: «*скр* *вд* Добре». Застосування подібного принципу забезпечує часову прив'язку кожної анотаційної позначки до конкретного місця у звуковому потоці, що є необхідним для примусового вирівнювання звуку та тексту (forced alignment) при тренуванні ASR.

Система розмітки мовних відхилень повинна бути чітко диференційована, має відповідати потребам соціолінгвістичного аналізу білінгвального мовлення та дозволяти ізолювати нелітературні форми при підготовці тренувального датасету для ASR. Маркер *р* позначає сегменти, вимовлені повністю російською мовою; маркер *сур* — суржикові вкраплення або окремі російські слова у переважно українському мовленні. Якщо в одному реченні вживаються і

українські, і російські слова, кожне російськомовне включення позначається окремо: «Я *сур* сегодня пішла *сур* домой». Якщо дитина неправильно вимовляє навіть саме російське слово, воно теж отримує позначку *сур*.

Під час розмітки аудіофайлів для тренування ASR важливо дотримуватися заздалегідь визначеної розмітки, адже вона дозволяє системі асоціювати конкретний звуковий фрагмент із правильним орфографічним відповідником. Маркери способу вимови відображають увесь спектр артикуляційних відхилень, характерних для дитячого мовлення: *роз* (розтягування звуку), *пскл* (по-складова вимова), *шеп* (шепелявість), *гарк* (гаркавість — неправильна вимова р/л), *спот* (спотворена вимова через складність звукового складу слова для дитини), *дслов* (неправильні словоформи — морфологічні помилки), *затр* (внутрішня затримка в слові), *об* (обмовка або самовиправлення). При цьому, залежно від типу відхилення, у текст вписується або літературний варіант слова (якщо зміст зрозумілий), або та його форма чи частина, яка реально прозвучала.

Хезитаційні маркери утворюють окрему розгалужену групу, що відображає когнітивно-просодичний вимір планування висловлювання, а також дозволяє позначити неробочі сегменти. До неї входять: *е* (наповнена пауза типу «е», «а», «и»), *мммм* (стан задумливості), *звроз* (розтягування звука в процесі пошуку слова), *пауз* (беззвучна пауза всередині сегменту), *м-м* (вагання), *вд* (вдих/видих). О. С. Іщенко досліджував паузування в усному мовленні і встановив, що його характер залежав від граматичного класу слова, після якого пауза виникає [Іщенко 2020, с. 45-55], а для дитячого мовлення ця залежність є ще більш вираженою через незавершеність лексичного вибору. Нарешті, маркери фонового шуму, а саме *шкор* (шум з коридору), *скр* (скрип), *стук*, *шар* (шарудіння папером), *крикд* (крик дитини у коридорі) тощо, необхідні для ідентифікації акустичних подій, не пов'язаних із мовленням, але здатних викликати збої алгоритмів розпізнавання [Bhardwaj, Othman, Kukreja, Belkhier, Bajaj, Srikanth Goud, Ur Rehman, Shafiq, Hamam 2022, с. 6].

На рівні «Ключові слова» слово фіксується у тій формі, як його реально вжила дитина, а поруч у дужках — нормативний літературний відповідник, якщо такий не було названо від початку. Якщо дитина не назвала ключове слово взагалі, сегмент залишається порожнім. Не можна применшити роль подвійної фіксації пари «фактично сказаного» і «передбаченого скриптом» у ключових словах, адже вона є критично важливою для порівняльного аналізу між тим, що дитина продукує, і тим, що система ASR мала б розпізнати.

Застосований підхід подвійного кодування лексем корелює із сучасними світовими стандартами корпусного опису усного спонтанного мовлення. Зокрема, на рівні спеціалізованого інструментария SpeCT 2.0, розробленого для програмного середовища Praat [Lennes 2019, с. 2–4], а також у загальній лінгвістичній практиці укладання усних корпусів, традиційно використовують принцип паралельних рядків (tiers) анотації. Цей метод передбачає структурне розведення інформаційних рівнів, тобто, наприклад, на одному з рядків фіксують безпосередньо вимовлене диктором (реалізований акустичний контур мовлення з усіма індивідуальними, діалектними чи віковими особливостями вимови), тоді як на паралельному рядку відображають нормативний літературний відповідник.

У випадку розробки спеціалізованого корпусу дитячого мовлення така форма двокомпонентного запису (правильні написання чи вимова і їхня фактична реалізація) є методологічно виправданою і дозволяє вирішити двовекторне завдання: з одного боку, зберегти автентичне дитяче мовлення для лінгвістичного аналізу, а з іншого — забезпечити систему автоматичного розпізнавання мовлення (ASR) верифікованими текстовими даними для навчання та оцінювання точності розпізнавання (сигнал — транскрипт — таргет). Проте варто зазначити, що, залежно від обставин, форма запису може бути дещо видозмінена з урахуванням відповідних потреб або особливостей опрацювання. Таким чином, поєднання фактично мовленого та когнітивно передбаченого скриптом результату можливе в межах як кількох суміжних рівнів анотації, так і на одному, що усе одно оптимізує процес подальшої комп'ютерної обробки отриманих корпусних даних.

2.3. Технологія інструментального аналізу та багаторівневої розмітки звукових сигналів у програмному середовищі Praat

Сучасний етап розвитку експериментальної фонетики та корпусної лінгвістики вимагає використання особливих і точних інструментальних засобів для верифікації акустичних властивостей усного мовлення. У процесі документування та лінгвістичного кодування спонтанного дискурсу, зокрема дитячого, першочерговим завданням є точна візуалізація, числове визначення і розрахунок параметрів звукового сигналу. Найбільш поширеним, гнучким та методологічно обґрунтованим інструментом для розв'язання цього кола завдань у світовій науковій практиці є спеціалізоване програмне середовище Praat.

Програмне забезпечення Praat (що в перекладі з нідерландської мови означає «говорити») є кросплатформним безкоштовним комп’ютерним комплексом, призначеним для проведення всебічного акустичного, фонетичного та фонологічного аналізу мовленнєвих сигналів. Свого часу цей інструментарій був розроблений на базі Амстердамського університету провідними фахівцями в галузі комп’ютерної фонетики Полом Бурсмою (Paul Boersma) та Давидом Вінінком (David Weenink). Станом на 2026 рік програма продовжує активно розвиватися, регулярно оновлюватися та адаптуватися до архітектурних вимог нових операційних систем, що підтверджує її високу актуальність та технічну стабільність як одного з провідних дослідницьких засобів сучасної комп’ютерної лінгвістики [Boersma, Weenink 2026, с. 341-347].

Програма Praat стала загальноприйнятим і обов'язковим інструментом у світовій лінгвістиці і є золотим стандартом для реалізації лінгвістичних проєктів, пов’язаних із побудовою мовленнєвих баз даних та тренуванням систем автоматичного розпізнавання тексту. Науковці та розробники систем розпізнавання мовлення (ASR) в усьому світі обирають Praat для анотування та аналізу звуку, оскільки його формати підтримуються більшістю сучасних лінгвістичних інструментів і бібліотек програмування. Робота, виконана в цій програмі, буде зрозумілою та придатною для інтеграції у будь-який міжнародний лінгвістичний проєкт без додаткової конвертації даних. Така затребуваність

зумовлена унікальним поєднанням аналітичного та прикладного функціоналу в одному інтерфейсі.

До основних базових можливостей програми, що безпосередньо використовуються під час кодування мовленнєвих сигналів, належать:

1. Відображення та редагування динамічної осцилограми;
2. Побудова спектрограм;
3. Автоматичне трекування та ручне коригування частоти основного тону;
4. Вимірювання формантних частот та інтенсивності;
5. Пакетне сегментування й створення баз анотованих даних.

Відображення та редагування динамічної осцилограми забезпечує наочне графічне представлення змін звукового тиску в часовому просторі, що дозволяє з високою точністю визначати межі мовленнєвих та шумових відрізків як базову умову для коректного розрізнення звуків і пауз. Цей процес доповнюється побудовою спектрограми, яка реалізує якісну візуалізацію частотного спектра звука, де по горизонталі відкладається час, по вертикалі — частота, а ступінь затемнення або колір об'єктивно відображує рівень інтенсивності енергії мовленнєвого сигналу.

Водночас автоматичне трекування та ручне коригування [F0] спрямоване на якнайточніший аналіз частоти основного тону — параметра, що безпосередньо корелює зі сприйняттям висоти людського голосу і фізично фіксує роботу голосових зв'язок мовця. Комплексний лінгвістичний аналіз передбачає також вимірювання формантних частот та інтенсивності шляхом математичного визначення локальних максимумів спектральної енергії ([F1], [F2], [F3]), які є основними акустичними інваріантами голосних звуків, паралельно з фіксацією загальної амплітуди звукового сигналу. Усе це, а також створення спеціальних текстових файлів-супутників, синхронізованих із часовою віссю вихідного аудіозапису для подальшої покрокової лінгвістичної розмітки тривалих записів, уможливлює процедуру пакетного сегментування й створення баз анотованих даних.

У своєму посібнику з використання програмного середовища Praat, У. Гут детально описує як архітектурну специфіку внутрішніх операцій роботи з програмою, так і прикладну методологію її використання саме в межах корпусної роботи з живим, непідготовленим спонтанним мовленням мовців різних вікових категорій [Gut 2008, с. 1-7]. У процесі аналізу мовлення молодших школярів функція візуального відображення спектрограми надзвичайно важлива, адже дитяче мовлення характеризується високим розташуванням формант та нестабільною артикуляцією, тож дослідник не може покладатися винятково на власний слуховий досвід. Зважаючи на це, можна безсумнівно сказати, що поєднання прослуховування аудіозапису та спектрографічного аналізу є необхідним і дозволяє досягти максимальної точної розмітки й об’єктивного лінгвістичного опису.

Центральним об'єктом лінгвістичної розмітки у Praat є файл TextGrid — структурований текстовий файл, що зберігається паралельно з відповідним аудіофайлом .wav і містить ієрархічно організовані рівні (tiers), прив'язані до точних часових позицій у секундах [Gut 2008, с. 1-7, Boersma, Weenink 2026, с. 341-347]. Кожен рівень може бути або інтервальним (interval tier), де анотується відрізок між двома часовими мітками, або точковим (point tier), де позначається один конкретний момент часу. Прив'язка лінгвістичного тегу до конкретного часового відрізку — принципова перевага Praat порівняно з системами суто текстової транскрипції, адже саме ця синхронізація є основою для подальшого примусового вирівнювання при тренуванні ASR [Radford, Kim, Xu, Brockman 2022, с. 1-4, Baevski, Zhou, Mohamed, Auli 2020, с. 1-8, Lennes 2019, с. 2].

Формат вхідного аудіофайлу є технічно значущим параметром. Для роботи із записами усного мовлення в Praat стандартом є нестиснений PCM WAV-файл, записаний з частотою дискретизації 16 кГц або вище та розрядністю 16 bit. Лише нестиснений формат гарантує точну взаємовідповідність між часовими мітками у файлі TextGrid та конкретними цифровими відліками аудіосигналу. Оскільки записи від початку переважно зберігаються у стисненому форматі .m4a, перед завантаженням у Praat кожен файл проходить конвертацію у формат .wav, що є

обов'язковим кроком під час технічної підготовки матеріалу. Пакет SpeCT 2.0 надає засоби часткової автоматизації рутинних операцій при роботі зі звуковими корпусами в Praat, зокрема — пакетного перейменування файлів, перевірки структури TextGrid та генерації базової статистики [Lennes 2019, с. 2-4].

Висновки до другого розділу

Другий розділ присвячено методиці та інструментальним засобам лінгвістичного анотування дитячого мовлення. Охарактеризовано емпіричну базу дослідження: 67 аудіозаписів (тривалістю 20–30 хвилин кожен), зібраних у квітні–травні 2024 року на базі Ліцею № 107 «Введенський» міста Києва у форматі індивідуальних напівструктурованих інтерв'ю. Умови проведення записів були стандартизовані, однак об'єктивно нездоланні зовнішні фактори шкільного середовища (фоновий шум, шкільний дзвінок, переривання) частково вплинули на акустичну якість окремих матеріалів.

До фінальної вибірки відібрано три записи учнів 2Б класу, а саме Віки К., Кіри Н. та Юхима К., які утворюють репрезентативну контрастивну тріаду за мовним профілем: переважно українськомовний, активно білінгвальний та фактично повністю російськомовний. Відбір здійснювався за критеріями якості звукозапису, повноти застосування методики та мовної однорідності соціолінгвістичного контексту. Методика інтерв'ювання, розроблена ГО «Спільномова» на основі міжнародних стандартів оцінки мовленнєвого розвитку (PPVT-5, EVT-3 і MAIN), включає п'ять послідовних тематичних блоків, що поєднують різні режими використання мови: від цілеспрямованого називання до розгорнутого монологічного опису.

Розроблено систему правил транскрибування та лінгвістичної розмітки. За основу прийнято норми Правопису 2019 р., а відхилення від них та інші явища фіксуються системою маркерів, обрамлених астерисками. Запропоновано чотирирівневу архітектуру TextGrid з паралельними інтервальними рівнями для мовлення модератора, мовлення дитини, ключових слів модератора та ключових слів дитини. Систему маркерів структуровано за п'ятьма функціональними групами: маркери мовних відхилень (*р*, *сур*); маркери способу вимови

(*роз*, *пскл*, *шеп*, *гарк*, *спот*, *дслов*, *затр*, *об*); хезитаційні маркери (*е*, *мммм*, *звроз*, *пауз*, *м-м*, *вд*); маркери паралінгвістичних явищ (*см*, *сміх*, *усм*, *хм*, *хв*); маркери фонового шуму (*шт*, *шкор*, *скр*, *стук*, *шар*, *крикд*). Принцип постановки маркерів безпосередньо перед відповідним сегментом забезпечує їхню часову прив'язку до звукового потоку, необхідну для примусового вирівнювання при тренуванні ASR.

Описано функціональну архітектуру програмного середовища Praat, а також роботу з файлами TextGrid і їхню перевагу порівняно з іншими засобами для транскрибації. Основними аналітичними функціями, задіяними у роботі, є: візуалізація осцилограми та широкосмугової спектрограми, трекування F_0 і вимірювання формантних частот, пакетне сегментування та автоматизована статистика через пакет SpeCT 2.0. Технічним стандартом вхідного сигналу визначено нестиснений PCM WAV з частотою дискретизації не нижче 16 кГц, що потребувало обов'язкової попередньої конвертації вихідних файлів .m4a.

РОЗДІЛ 3. ПРАКТИЧНА РЕАЛІЗАЦІЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ ЛІНГВІСТИЧНОГО АНОТУВАННЯ ЗАПИСІВ

Третій розділ зосереджений на практичному складнику дослідження й описує фактичну реалізацію засвоєних теоретичних засад і попередньо розробленої методики безпосередньо на матеріалі відібраних аудіозаписів. Його мета — не лише описати виконані операції, а й проаналізувати їхні результати, представити характеристики сформованого корпусу та оцінити його потенціал для донавчання моделей автоматичного розпізнавання мовлення.

Розділ складається з трьох підрозділів. У підрозділі 3.1 описано підготовчий технічний етап і процедуру мікросегментації та текстової транскрибації аудіофайлів у середовищі Praat. Підрозділ 3.2 зосереджено на специфічних явищах дитячого мовлення й анотаційних рішеннях, яких вони вимагають. У підрозділі 3.3 здійснено кількісний і якісний аналіз характеристик сформованого корпусу, а саме лексичних показників, частотності маркерів та акустико-фонетичного профілю мовців.

3.1. Процедура сегментації та текстової транскрибації аудіофайлів

Емпіричний матеріал, отриманий у ході польових досліджень у Ліцеї № 107 «Введенський» м. Києва, первинно надійшов у форматі .m4a. Параметри запису, здійсненого за допомогою вбудованих цифрових диктофонів мобільних пристроїв на базі операційної системи Android, характеризувалися частотою дискретизації 44 100 Гц та величиною бітрейту 128 кбіт/с.

Конвертація аудіозаписів була здійснена за допомогою хмарної веб-платформи Convertio [Convertio 2026] відповідно до вимог, описаних у підрозділі 2.3. Процедура включала завантаження вихідних файлів .m4a на сервер через веб-інтерфейс сервісу, вибір цільового формату розширення (WAV) та налаштування параметрів кодування в додатковому меню (вибір аудіокодека Signed 16-bit PCM та збереження частоти дискретизації на рівні 44 100 Гц). Після завершення процесу віддаленої пакетної декомпресії готові файли були завантажені на локальний диск для подальшої роботи (рис. 3.1). У результаті технічної конверсії було сформовано три робочі звукові файли. Верифікація

коректності перекодування здійснювалася шляхом контрольного відтворення файлів у медіаплеєрі, розробленому компанією Microsoft, а також зіставлення тривалості аудіодоріжок до і після процедури онлайн-конвертації, що дозволило повністю виключити ймовірність виникнення технічних розривів, випадіння фреймів чи часових зсувів.

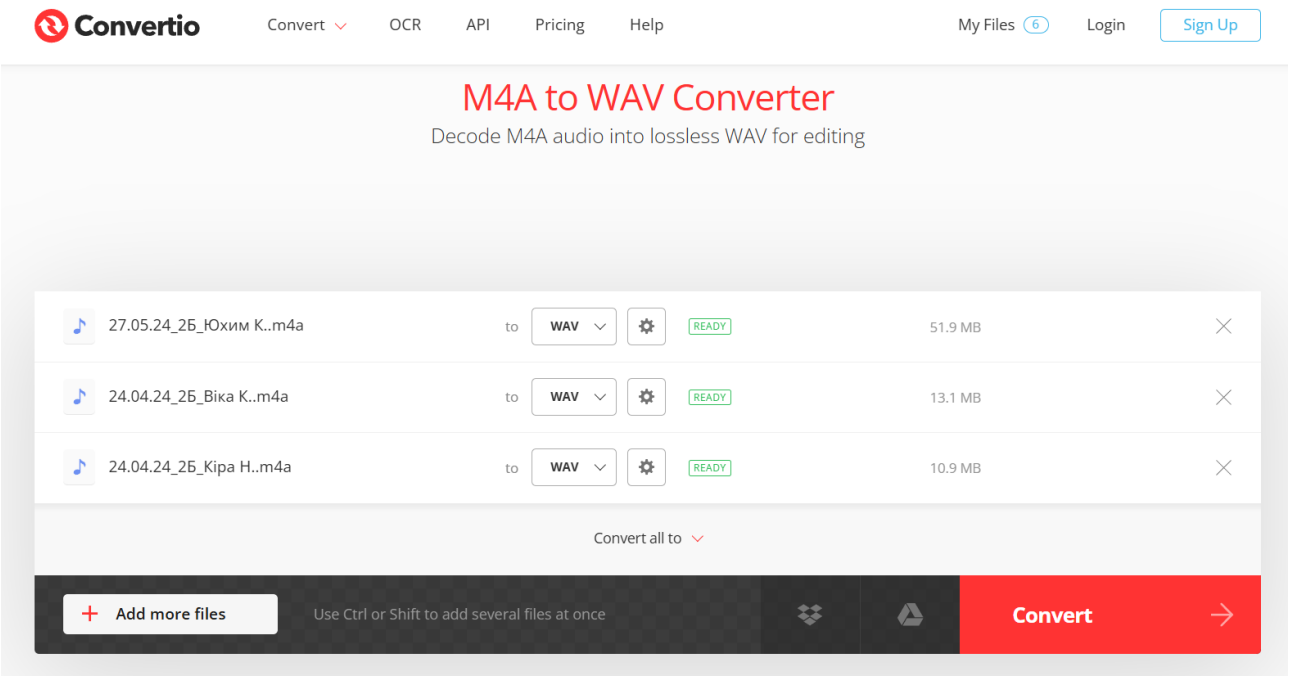

Рисунок 3.1

Наступний етап практичної роботи передбачав імпорт одержаних WAV-файлів у спеціалізоване програмне середовище Praat за допомогою базової команди «Read from file». Первинний аналіз імпортованого матеріалу полягав у загальному візуальному огляді сигналу у комбінованому вікні «Sound + Spectrogram» з використанням інструментів та налаштувань, описаних у підрозділі 2.3. Налаштування спектрального аналізу передбачали встановлення верхньої межі частоти до 5000 Гц. Уже на цьому первинному етапі візуального та слухового моніторингу фіксувались загальні особливості запису, наприклад, наявність тривалих ділянок із фоновим шумом коридору чи моменти раптових підвищень рівня сигналу через пересування меблів, а також виявлення основних елементів структури інтерв'ю й зміна мовців відповідно до них. Паралельно здійснювалось повторне прослуховування для загальної орієнтації в структурі

інтерв'ю, виявлення ключових точок зміни мовців, а також оцінки загальної розбірливості записаного матеріалу.

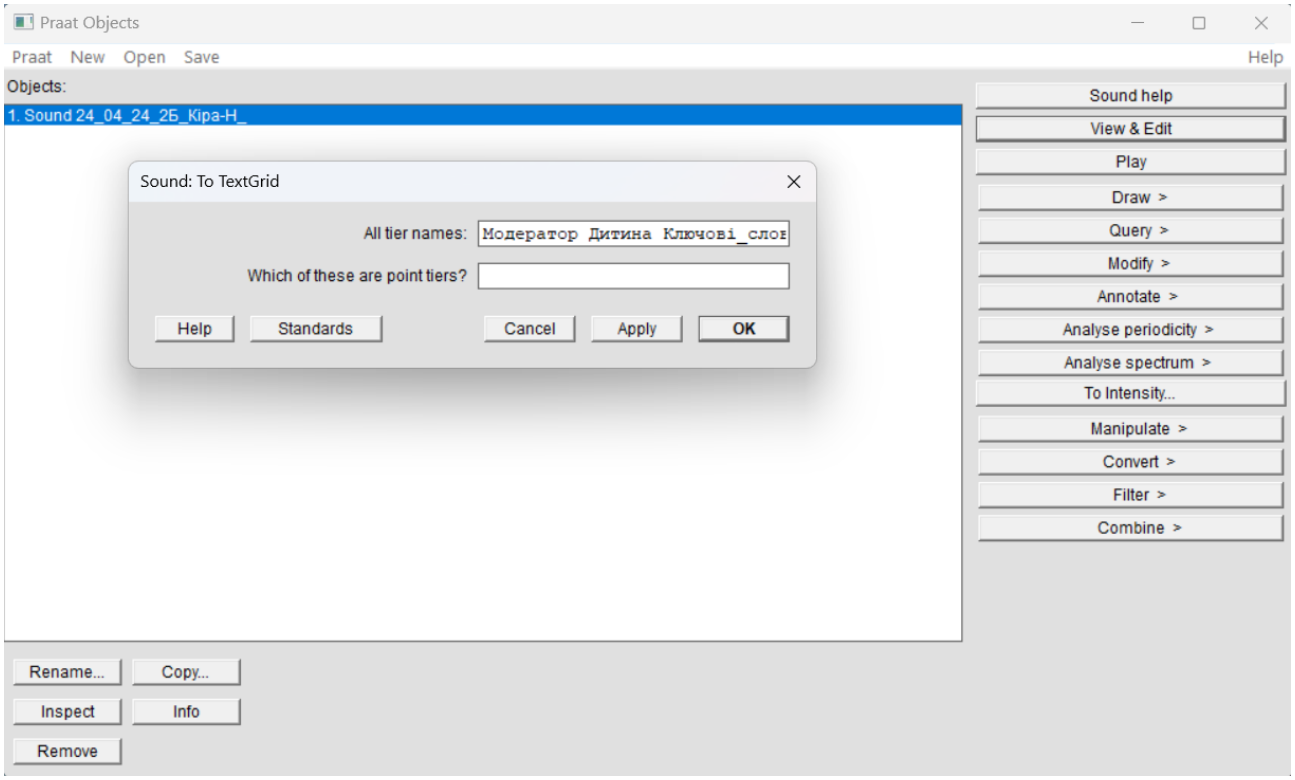

Рисунок 3.2

Для безпосереднього проведення багаторівневого лінгвістичного анотування для кожного WAV-файлу засобами меню «Annotate» та «To TextGrid» створювався однойменний синхронізований файл розмітки TextGrid. Відповідно до розробленої у Розділі 2 структури TextGrid, було створено чотири паралельні інтервальні рівні з відповідними назвами («Модератор», «Дитина», «Ключові слова (М)», «Ключові слова (Д)») (рис. 3.2).

Водночас, відповідно до практичних умов транскрибування, розмітка здійснювалася дискретно, тобто у файлі TextGrid виокремлювалися й наповнювалися текстовим вмістом лише ті сегменти, де об'єктивно було наявне говоріння чи звукова активність мовців. Періоди взаємного мовчання, тривалі фізіологічні паузи та сторонні фонові шуми залишалися нерозміченими у вигляді порожніх інтервалів (empty intervals). Так, наприклад, у одному з фрагментів відповідним маркером (*шт*) було позначено помірно зашумлені періоди мовленнєвої активності модератора й дитини, окремо виділено момент хезитації

в процесі міркування (*мммм*), а також промарковано неробочий сегмент із криком дітей у коридорі (*крикд*), який через його інтенсивність система могла б помилково сприйняти як те, що відбувається між модератором і опитуваною дитиною (рис. 3.3). Завдяки такому підходу корисний звуковий сигнал було ізольовано від пауз, що запобігає помилковому спрацьовуванню моделі на ділянках тиші чи стороннього шуму [Radford, Kim, Xu, Brockman 2022, с. 1-4].

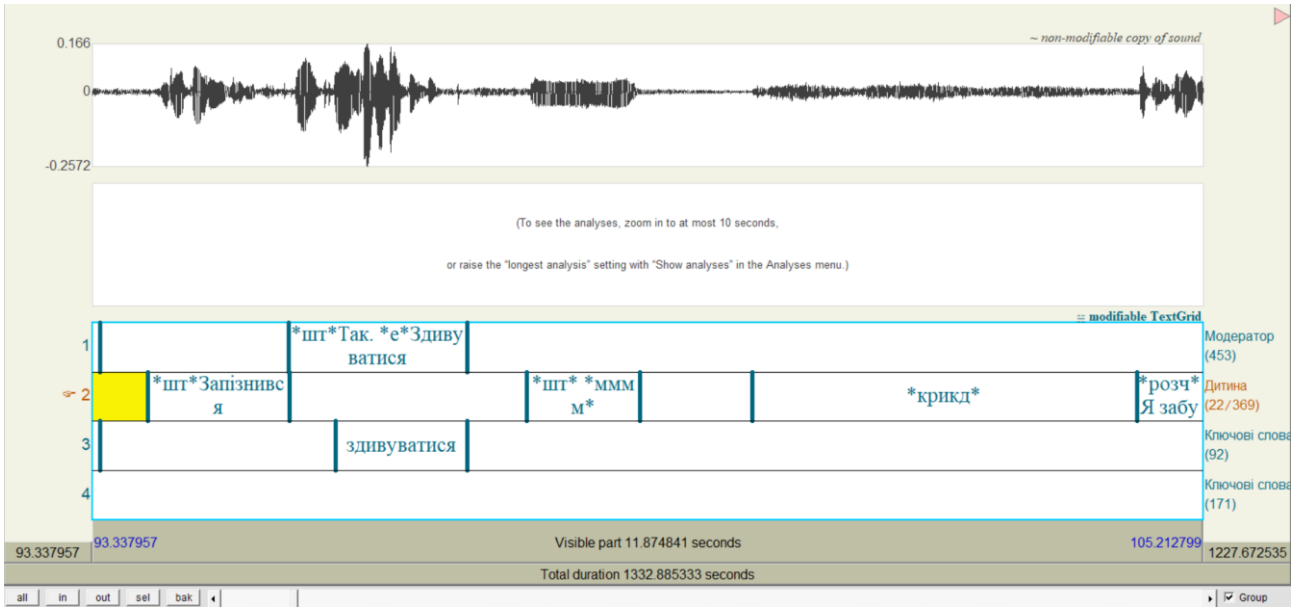

Рисунок 3.2

Процедура мікросегментації здійснювалася вручну в інтерактивному редакторі «Sound + TextGrid Editor». Встановлення часових меж суміжних інтервалів виконувалося на основі поєднання акустичного та спектрального критеріїв, описаних у підрозділі 2.3. Складну аналітичну проблему становили зони паралельного або накладеного мовлення мовців (позначалися маркером *син*, *дит*, *мд*). У таких фрагментах часові межі інтервалів визначалися максимально точно на слух і встановлювалися індивідуально на відповідних рівнях («Модератор» і «Дитина»), оскільки реальний початок та затухання звукових хвиль мовців переважно не збігалися в часі. Так, наприклад, наприкінці репліки модератора «Добре» вже чути початок відповіді дитини «Похоже на тукана», при цьому мовлення модератора усе ще триває, а потім разом із емоційною реакцією розпочинається ще раз, і водночас дитина ще вимовляє останнє слово (рис. 3.4).

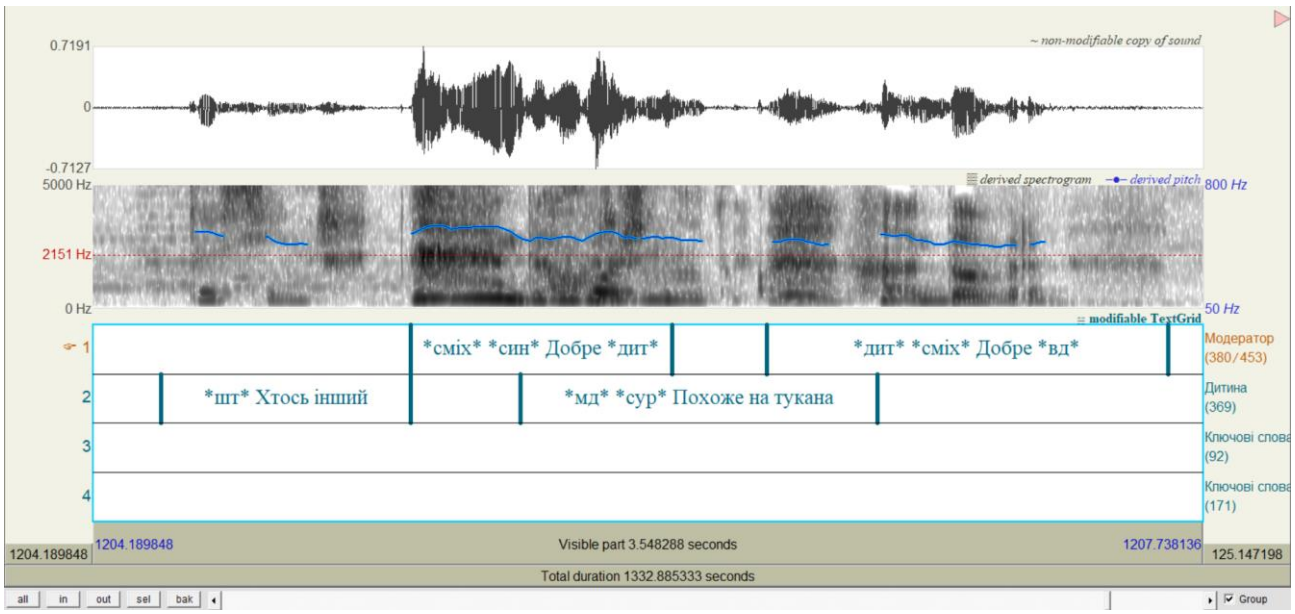

Рисунок 3.3

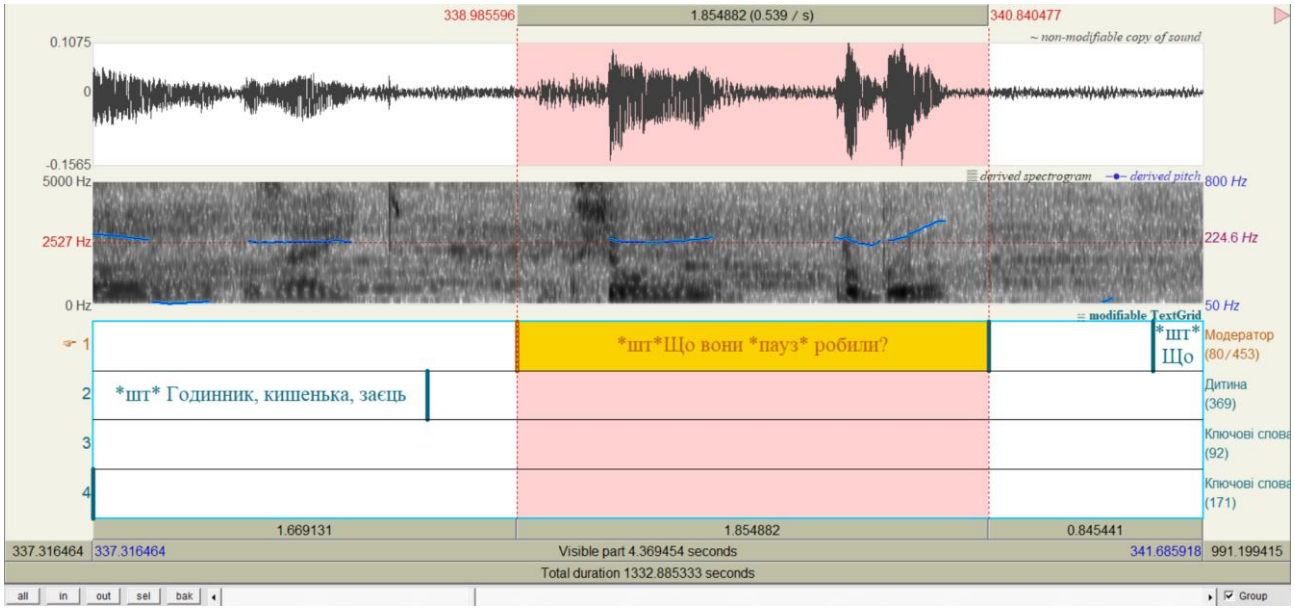

Рисунок 3.5

Мінімальною одиницею сегментації було визначено інтонаційно-комунікативне висловлювання одного мовця, що не переривається тривалими паузами. Паузи всередині мовленнєвого потоку мовця, тривалість яких не перевищувала 0,5 с, розглядалися як несегментоутворюючі й маркувалися тегом *пауз* безпосередньо всередині текстового поля поточного інтервалу. Натомість незаповнені паузи, що тривали понад 0,5 с, інтерпретувалися як значущі межі комунікативних реплік або сигнали завершення висловлювання, що вимагало виставлення нової часової межі та створення наступного інтервалу. Так,

наприклад, інтервал між «що вони» і «робили» становить 0,4 с, а між цією реплікою і наступною «Що відбувалося» інтервал становить 0,6 с (рис. 3.5).

Текстова фіксація мовленнєвого матеріалу в межах виділених сегментів була виконана за нормами чинного правопису (редакція 2019 року). У разі спотворення слова вписувалася фактично вимовлена одиниця з відповідним маркером із системи розмітки, визначеної у підрозділі 2.2 (*спот*, *сур*, *дслов*). Наприклад, у реченні «Ми катались на човні, і нас облила вода» слово «човен» вжите із неправильним закінченням («човну», тоді як нормативна форма місцевого відмінка однини — «човні»). З урахуванням того, що це слово входить також і до переліку ключових, відповідні порушення у вимові було відображено і на четвертому рівні «Ключові слова (Д)» (рис. 3.6).

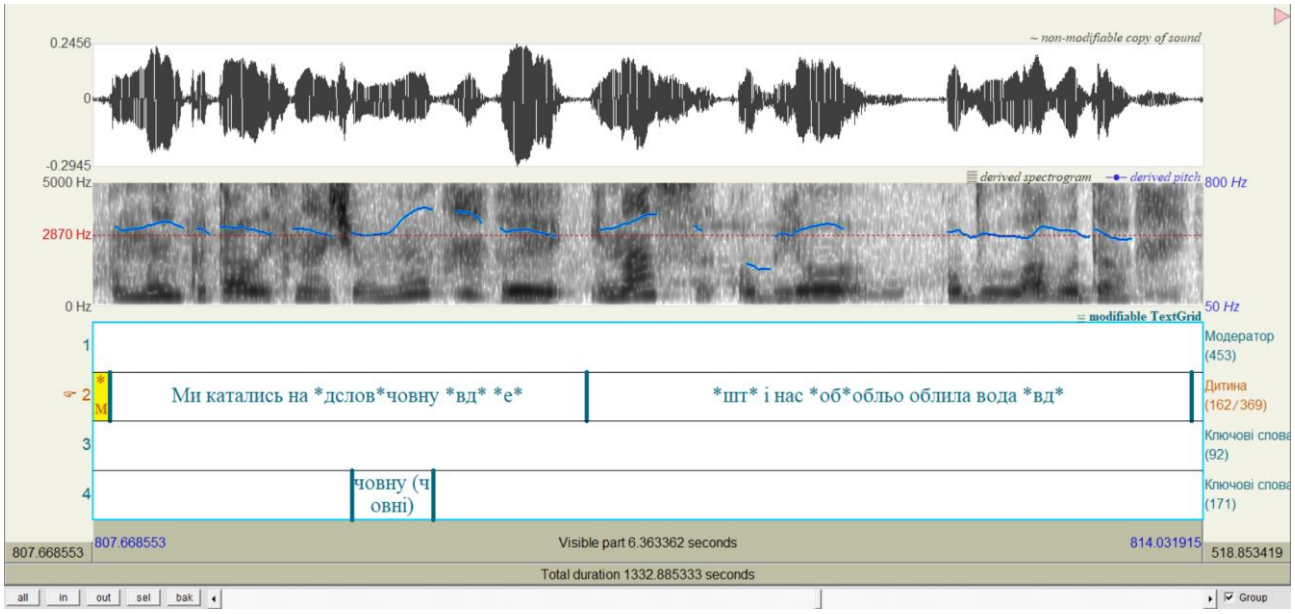

Рисунок 3.4

У випадку сегментів із надмірно поганою розбірливістю, замість спроби реконструювати слово, перевага надавалася позначці *нерозб*, оскільки неправильно анотований сегмент є шкідливішим для тренування ASR-системи, ніж пропущений [Gerosa1, Giuliani, Narayanan, Potamianos 2009, с. 4-6]. Наприклад, у момент, коли дитина не до кінця розуміє завдання, вона втрачає впевненість у своїх словах і вимовляє їх тихо чи взагалі починає бубоніти. Показники осцилограми чи спектрограми у цьому випадку дозволяють лише визначити межі чітко вимовленого слова «хмаринка», але є недостатніми, аби

визначити зміст решти фрази. Якщо сприймати цей момент запису винятково акустично, то здається, ніби мовець послуговується ненормативною лексикою, що неможливо з точки зору логіки й аналізу поведінки дитини протягом усього іншого часу запису. Оскільки однозначно визначити сказане неможливо, у стенограмі було вказано лише точно почуте слово «хмаринка», а решта промаркована як нерозбірливе мовлення (рис. 3.7).

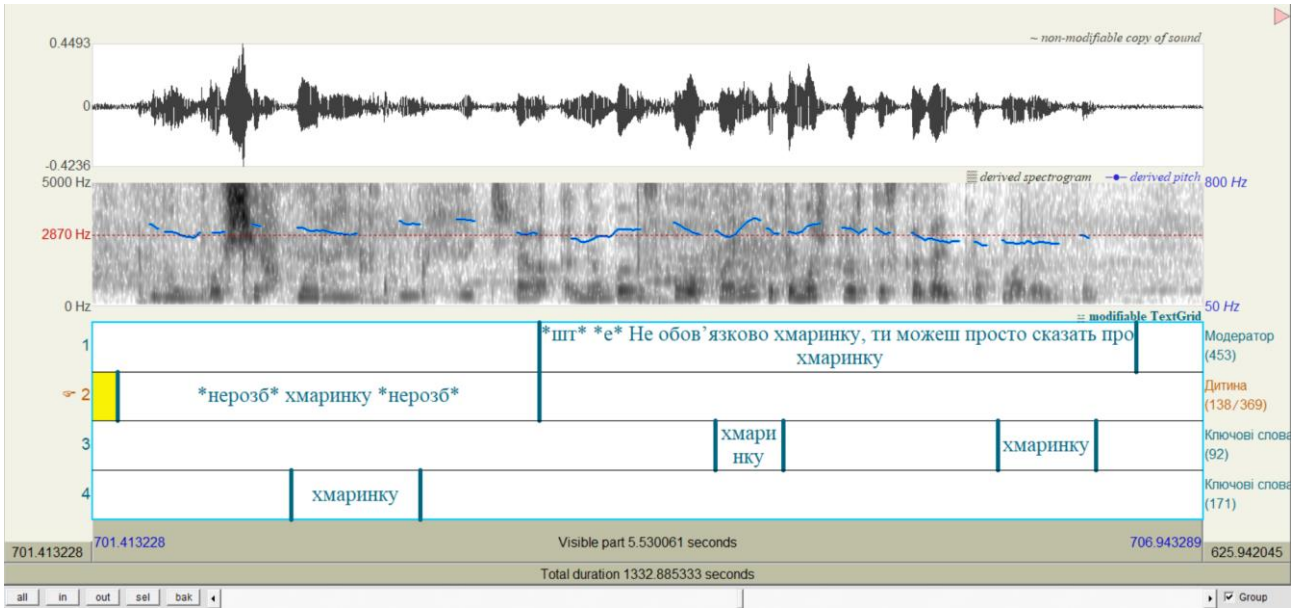

Рисунок 3.5

Після завершення первинного анотування було повторно прослухано весь файл від початку до кінця, текст звірено зі звуком і виправлено неточності. Основними типами помилок були:

1. неточно виставлені межі мовленнєвої активності;
2. орфографічні й друкарські помилки;
3. пропущені чи неправильно визначені маркери.

Після верифікації файл TextGrid зберігається у тому самому каталозі, що і відповідний аудіофайл у форматі .wav, із автоматично наданим алгоритмами програми Praat ідентичним ім'ям файлу для подальшої зручності пошуку й паралельного використання (рис. 3.8).

Загальні часові витрати на повне лінгвістичне анотування та розмітку одного звукового запису (без урахування попередньої декомпресії та конвертації форматів) становили від 11 до 15 астрономічних годин, що вказує на високу

трудомісткість корпусних досліджень спонтанного мовлення. Найбільше часу знадобилося для аналізу запису мовлення Юхима К. через високу частоту кодових переключень та необхідністю ретельної ідентифікації меж мовних переходів для кожного мікросегмента. Натомість записи, зроблені за участі Віки К. і Кіри Н. характеризувався вищою швидкістю обробки завдяки чіткішому мовленню, стабільному україномовному контуру та нижчому рівню шуму. Одержані розбіжності в часових витратах є об'єктивним якісним показником лінгвістичної складності матеріалу, що доводить, що саме глибина та точність розмітки, а не лінійний обсяг сирих даних, є визначальним фактором при створенні навчальних вибірок для сучасних ASR-систем.

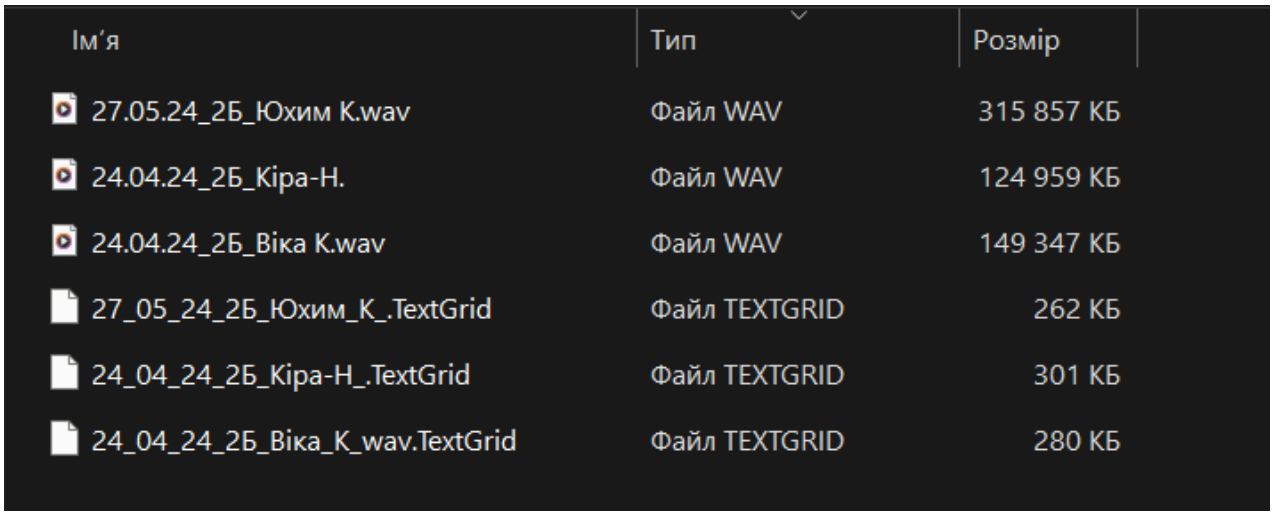

Рисунок 3.6

3.2. Особливості розмітки специфічних явищ дитячого мовлення

Специфіка мовлення молодших школярів зумовлює систематичну появу цілого ряду явищ, що не мають аналогів у мовленні підготовлених дорослих дикторів і вимагають особливих анотаційних рішень. Саме ці явища формують головну проблему для систем автоматичного розпізнавання при роботі з дитячим мовленням [Gerosa1, Giuliani, Narayanan, Potamianos 2009, с. 4-6, Bhardwaj, Othman, Kukreja, Belkhier, Bajaj, Srikanth Goud, Ur Rehman, Shafiq, Hamam 2022, с. 2-6].

Хезитаційні явища та паузації — найчастотніша категорія в аналізованому матеріалі. Відповідно до класифікації, наведеної у підрозділі 2.2, вони

фіксувалися тегами *пауз*, *е*, *ммммм*, *звроз* тощо. Практичний аналіз підтвердив тенденцію, описану у підрозділі 2.2: у дітей порівняно з дорослими мовцями затримки лінійного розгортання фрази безпосередньо пов'язані з когнітивним навантаженням під час вибору повнозначних слів. Як було встановлено у підрозділі 2.2, хезитаційні паузи у дітей є ще частотнішими, ніж у дорослих, і з’являються у визначеному порядку. Так, наприклад, у мовленні учнів Ліцею № 107 «Введенський» найчастіше паузи виникають поруч із дієсловами й іменниками, при цьому можуть бути доповнені диханням (*вд*) чи хезитаціями (*хм*) (рис. 3.9).

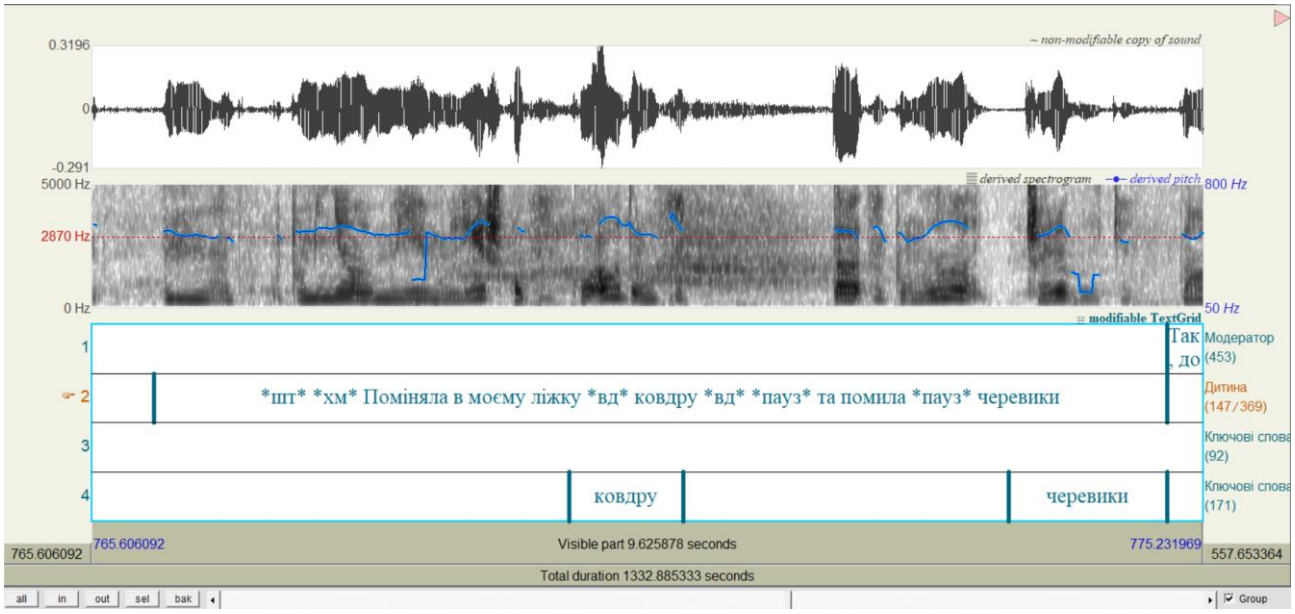

Рисунок 3.7

Повтор є поширеним явищем у дитячому мовленні, дитина може декілька разів розпочинати той самий відрізок, доки не знайде потрібну форму. Іноді діти вдаються до самокорекції, коли помічають власну помилку і виправляють її. Потім, за наявності, прописувався виправлений варіант слова. Подібні явища є прямим свідченням процесу засвоєння мови й правил її функціонування, який ще триває у дітей, а також того, що дитина активно навчається не лише коли її виправляють інші (наприклад, батьки чи вчителі), а й вчиться самостійно контролювати власне мовлення. Кожну таку обмовку було зафіксовано таким способом, як прозвучало слово (тобто помилкова форма до виправлення), а перед цим проставлено маркер *об* (рис. 3.10).

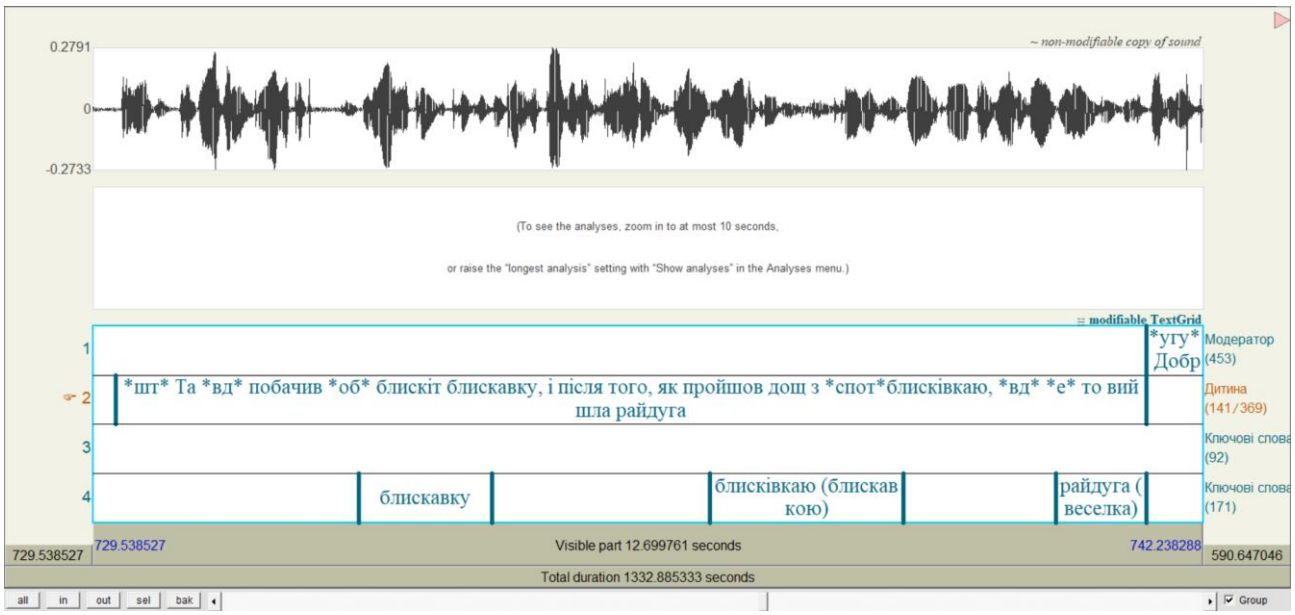

Рисунок 3.8

Епізоди порушення черги говоріння та емоційні реакції маркувалися тегами *мд* чи *дит*, залежно від домінантного мовця у сегменті. Наприклад, у фрагменті, де дитині показують картки, кожна нова картинка викликає у дитини сильну емоцію подиву й захоплення, через що школяр час від часу вигукує «о» (рис. 3.11). Позитивним аспектом наявності таких вигуків є можливість зрозуміти, як під час проведення опитування, так і потім під час аналізу, інтерес дитини до завдання й активну залученість у процес.

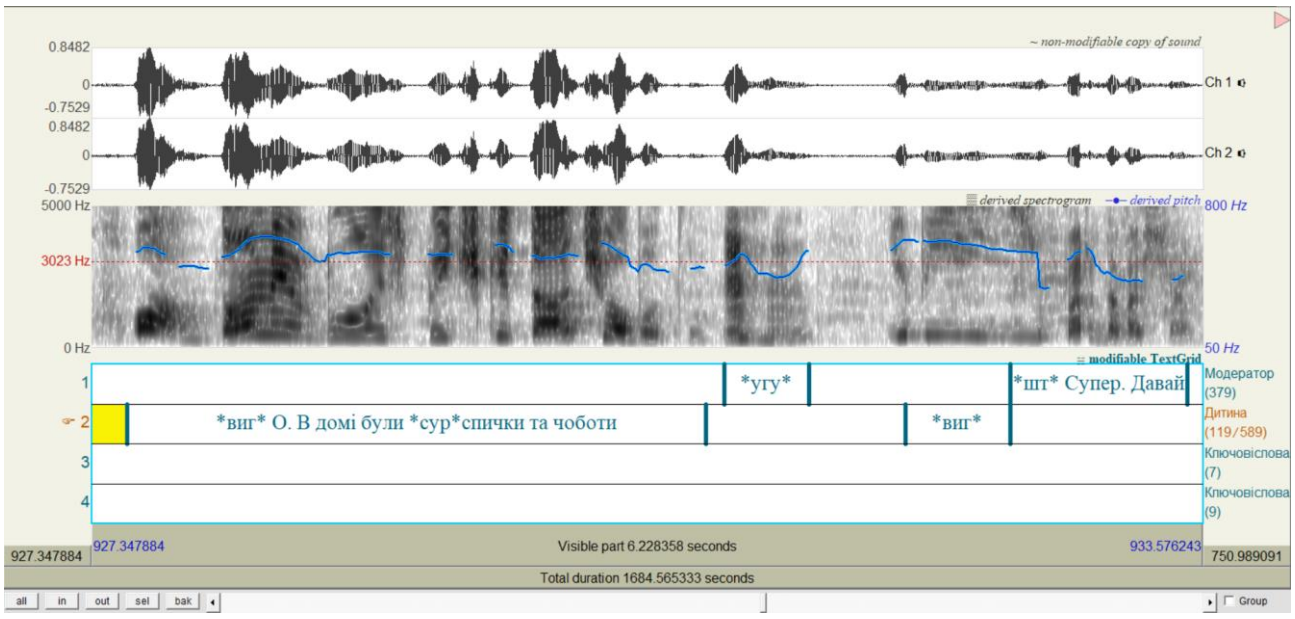

Рисунок 3.9

Дитяче мовлення під час інтерв'ю насичене невербальними, паралінгвістичними проявами, що супроводжують чи повністю замінюють вербальну відповідь. Відповідно до класифікації, наведеної у підрозділі 2.2, вони фіксувалися тегами *см* (одиночний сміх), *смд* (сміх із вдохом або видохом), *сміх* (сміх протягом цілого сегменту), *усм* (усміхнена вимова), *хм* (хмикання), *хв* (хвилювання в голосі), *усвід* (реакція усвідомлення) тощо на місці відповідного звука. Наприклад, на етапі роботи з картками, де дитина повинна скласти речення, вона отримує картки з лексемами «годинник», «літак» і «кавун». Оскільки, за умовами завдання, складені речення можуть бути як реалістичними за змістом, так і абсолютно абсурдними чи фантастичними, таке поєднання карток смішить дитину, тому спочатку мовець сміється (паралінгвістичний сигнал), а потім ословлює свої враження зі стверджувальною інтонацією: «Не, смешно, правда» (рис. 3.12).

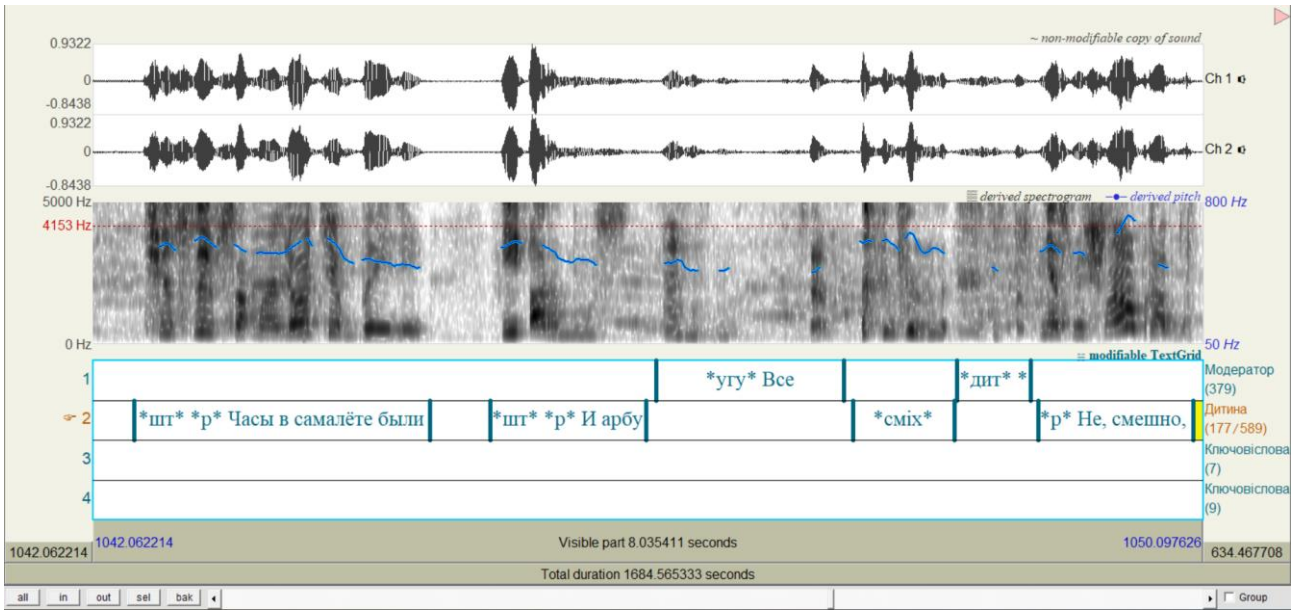

Рисунок 3.10

3.3. Кількісні та якісні характеристики сформованого корпусу

До аналізованого корпусу включено три записи загальною тривалістю 76 хвилин 49 секунд. Тривалість кожного окремого запису становить від 20 до 30 хвилин і відповідає повному циклу методики. Після завершення повного анотування оцінюються такі кількісні параметри:

1. загальна кількість слів на рівні «мовлення дитини»;

2. кількість виокремлених ключових слів;
3. кількість інтервалів із позначками нерозбірливості;
4. кількість хезитаційних маркерів;
5. кількість маркерів зміни мови;
6. найчастотніший маркер;
7. частота основного тону.

Оцінка статистичних параметрів TextGrid-об'єктів здійснювалася за допомогою заздалегідь розроблених скриптів автоматизованого підрахунку елементів для середовища Praat, а також шляхом покрокового ручного моніторингу текстових підписів інтервалів розмітки (рис. 3.13).

```
intervals [176]:
    xmin = 1048.903521932991
    xmax = 1050.025201973241
    text = "*р* Не, смешно, правда"
intervals [177]:
    xmin = 1050.025201973241
    xmax = 1058.5798608273078
    text = ""
intervals [178]:
    xmin = 1058.5798608273078
    xmax = 1061.2483526057067
    text = "Отакі *сур* кроссовочки пацанские"
intervals [179]:
    xmin = 1061.2483526057067
    xmax = 1068.7817101647247
    text = ""
intervals [180]:
    xmin = 1068.7817101647247
    xmax = 1076.6635477529885
    text = "*шт* *р* На, чтоб пойти на площадку надо снять тапки и положить конфету"
intervals [181]:
    xmin = 1076.6635477529885
    xmax = 1077.088497833783
    text = ""
intervals [182]:
    xmin = 1077.088497833783
    xmax = 1080.3393639005503
    text = "*шт* *р* И *йой* не бросать конфетку в песок"
intervals [183]:
```

Рисунок 3.11

Для унаочнення загального масштабу дослідження, первинної структуризації емпіричного матеріалу та проведення компаративного аналізу між мовленнєвими профілями трьох інформантів нижче наведено таблицю, де відображено базові кількісні та структурні метрики корпусу (табл. 3.1).

Аналіз кількісних параметрів, відображених у таблиці (табл. 3.1), демонструє суттєві соціолінгвістичні та прагматичні розбіжності в мовленнєвій поведінці дітей, попри повну стандартизованість умов проведення експерименту. Загальний лексичний обсяг дитячих реплік у корпусі становить 1534 слова (токени).

Таблиця 3.1

Аналізований параметр корпусу	**Віка К.**	**Кіра Н.**	**Юхим К.**	**Разом**
Індивідуальна тривалість запису (хв, с)	26:33	22:12	28:04	**76:49**
Загальна кількість слів (рівень «Дитина»)	566	511	457	**1534**
Кількість названих ключових слів (рівень «Дитина»)	66	85	52	**203**
Середня частота основного тону голосу F_0 (Гц)	297.6	240.6	277.6	**271.9**

Серед опитаних дітей найвищу абсолютну мовленнєву продуктивність виявила Віка К., яка сказала 566 слів за 26 хв 33 с. Водночас найвища дискурсивна щільність (темп вербалізації) властива Кірі Н., яка за найкоротшу за часом розмову (22 хв 12 с) вимовила 511 слів, що свідчить про високу швидкість лінгвального планування та розвинені навички спонтанного монологічного мовлення. Натомість Юхим К. продемонстрував найнижчу продуктивність — лише 457 слів за найдовшу за часом сесію (28 хв 04 с), що вказує на складнощі в комунікації, уповільнений темп мовлення та наявність бар'єрів при продукуванні висловлювань, або ж такі результати можуть бути зумовлені неточностями, що могли виникнути в процесі анотації.

Цікава закономірність виявляється при зіставленні загального обсягу слововживань із кількістю успішно названих ключових слів на рівні «Ключові слова (Д)». Кіра Н. випродукувала найбільшу кількість цільових лексем — 85 одиниць при загальному обсязі мовлення у 511 слів. Це вказує на високу концентрацію уваги дитини на предметних ілюстраціях та її прагнення точно слідувати вказівкам модератора. У Юхима К. зафіксовано найменшу кількість

ключових слів (52), що безпосередньо корелює із загальним зниженням обсягу його вербалізації та частим уникненням номінації складних об'єктів.

Друга найважливіша координата аналізу сформованого корпусу пов'язана з описом частотного розподілу метатекстових маркерів, лінгвістичних тегів відхилень та елементів акустичного фону. На відміну від суто текстових корпусів, у розмовному датасеті метамаркери розподіляються по всій тривалості звукової хвилі і фіксуються повнозначно незалежно від того, хто виступає їх безпосереднім джерелом — дитина, дослідник чи навколишнє середовище. Зведена інформація щодо частотності використання анотаційних тегів відображена у таблиці (табл. 3.2).

Таблиця 3.2

Тип анотаційного маркера	**Запис Віки К.**	**Запис Кіри Н.**	**Запис Юхима К.**	**Сумарно в корпусі**
Маркер шуму тла (*шт*)	153	204	54	**411**
Заповнені вокалізовані паузи хезитації (*е*)	49	38	15	**102**
Внутрішньосегментні незаповнені паузи (*пауз*)	6	50	1	**57**
Лексичні обмовки та самокорекції (*об*)	3	14	2	**19**
Повністю російськомовні сегменти (*р*)	8	0	49	**57**
Прояви інтерференції та змішаного мовлення (*сур*)	10	36	20	**66**
Маркери абсолютної нерозбірливості сигналу (*нерозб*)	3	10	2	**15**
Модероване мовлення (*мод*)	19	19	45	**83**

Аналіз даних частотного розподілу, наведених у таблиці (табл. 3.2), дозволяє визначити якісну та кількісну домінанту сформованого корпусу: найчастотнішим абсолютно усюди виявився маркер недиференційованого шуму тла *шт*. У підсумку цей тег зафіксовано 411 разів (153 рази у записі Віки К., 204 рази у Кіри Н. та 54 рази у Юхима К.). Така значна представленість шуму зумовлена польовим характером збирання емпіричного матеріалу (безпосередньо у приміщеннях діючого загальноосвітнього ліцею під час навчального процесу), що неминуче сприяло наповненню звукової доріжки екзогенними акустичними явищами, переважно шумом кроків і криками з коридору чи спортзалу, а також скрипом шкільних меблів чи шарудінням паперу під час зміни положення робочих матеріалів.

З точки зору розробки та донавчання систем автоматичного розпізнавання мовлення (ASR), така повсюдна присутність тегу *шт* не є технічним недоліком, а сприяє покращенню роботи моделі. Промислові моделі розпізнавання, навчені на ідеальних стерильних аудіокнигах чи студійних подкастах дорослих дикторів, виявляються абсолютно безпорадними у реальних умовах експлуатації через ефект акустичного розходження (acoustic mismatch). Тож наявність розмічених сегментів із тегом *шт* дозволяє під час процедури fine-tuning адаптувати нейромережу, навчивши її успішно ігнорувати немовленнєві акустичні події і не інтерпретувати їх як помилкові фонеми чи лексеми, що суттєво знижує показник похибки Word Error Rate (WER).

Детальний якісний аналіз лінгвістичних маркерів виявляє глибокі внутрішньосистемні відмінності між когнітивними та соціолінгвістичними стратегіями мовців. Група хезитаційних маркерів (*е*, *пауз*) відіграє роль індикатора швидкості мовленнєвого планування. У Віки К. зафіксовано 49 заповнених пауз (*е*) при мінімальній кількості незаповнених зупинок (6) та вкрай низькому рівні обмовок (3). Це свідчить про те, що дитина підтримує високий темп мовлення, оперативно заповнюючи мовленнєві паузи вокалізаціями, не зупиняючи лінійне розгортання фрази і майже не здійснюючи рестартів синтаксичних контурів.

Діаметрально протилежну картину демонструє запис Кіри Н. Тут спостерігається виражений каскадний характер когнітивного навантаження: 38 заповнених пауз (*е*) супроводжуються 50 тривалими внутрішньофразовими незаповненими паузами (*пауз*) та найвищим у корпусі рівнем лексичних обмовок і самокорекцій (*об*) (14 фіксацій). Кіра Н. активно конструює складні синтагми (що доводиться рекордною кількістю названих ключових слів — 85), проте цей процес вимагає безперервного внутрішнього моніторингу, викликає моторні збої, обриви слів та свідомі граматичні заміни безпосередньо в під час мовлення. Висока частота інтервалів абсолютної нерозбірливості *нерозб* (10) у її записі також пов'язана із моментами сильного падіння інтенсивності голосу наприкінці важких фразових конструкцій.

Соціолінгвістичний вимір корпусу, обумовлений специфікою двомовного середовища, найчіткіше виражений через диференціацію маркерів повністю російського мовлення *р* та суржикових елементів *сур*. У записі Кіри Н. повністю відсутні суцільні російські фрази (*р* = 0), проте зафіксовано високий рівень суржикової інтерференції (*сур* = 36). Це свідчить про функціонування стійкої, але змішаної мовної системи на базі української мови інформанта.

Найбільш складну якісну картину має профіль Юхима К. Його мовлення характеризується тотальним домінуванням російської — зафіксовано 49 великих інтервалів із тегом *р*, які поєднуються із 20 суржиковими вкрапленнями.

У нього незначна кількість хезитацій (*е* = 15, *пауз* = 1) та самокорекцій форм (*об* = 2). Замість цього Юхим К. використовує стратегію комунікативної пасивності, що викликало дзеркальне зростання активності дослідника-інтерв'юера. Тег *мод* (інтенсифікація реплік, навідних запитань та прямих лінгвістичних підказок модератора) у записі Юхима К. досягає рекордних 45 фіксацій (тоді як у Віки та Кіри — лише по 19) й найбільше виявлена на останньому етапі перевірки неправильно (або російською) названих карток.

Четвертий фундаментальний вимір аналізу сформованого корпусу — акустико-фонетичний профіль мовців, виражений через вимірювання частоти основного тону голосу (F_0). Послідовності математичних значень відображають

індивідуальні виміри F_0 в Герцах, зроблені в середовищі Praat шляхом визначення частоти основного тону на кількох сегментах тривалістю до 10 секунд на приблизно однакових відстанях впродовж усього запису й подальшому розрахунку середнього арифметичного:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Послуговуючись наведеною формулою, можемо розрахувати наступне:

Віка К.: 289 Гц + 293 Гц + 310 Гц + 285 Гц + 311 Гц. Середнє арифметичне значення становить 297.6 Гц.

Кіра Н.: 239 Гц + 241 Гц + 241 Гц + 246 Гц + 236 Гц. Середнє арифметичне значення становить 240.6 Гц.

Юхим К.: 231 Гц + 243 Гц + 322 Гц + 280 Гц + 312 Гц. Середнє арифметичне значення становить 277.6 Гц.

Зважаючи на обмежений обсяг вимірювальних значень (5 сегментів для кожного мовця), отримані показники середньої частоти основного тону мають винятково ілюстративний характер і слугують винятково для демонстрації загального частотного контуру дитячих голосів у межах цього проєкту, а тому не претендують на вичерпну статистичну повноту.

Середньозважений показник частоти основного тону по всьому сформованому корпусу становить 271.9 Гц. Для дорослого чоловічого голосу нормативний діапазон F_0 становить 100–150 Гц, для жіночого — 180–220 Гц. Показники другокласників ліцею, значно вищі за 250 Гц, наочно демонструють природу повної несумісності стандартних комерційних акустичних моделей із дитячим мовленням. Анатомічно короткі голосові зв'язки та малий об'єм резонаторного тракту дітей зміщують увесь спектральний контур і формантні частоти вгору.

При цьому якісний аналіз динаміки F_0 виявляє високу стабільність вокального регістру Кіри Н. (відхилення в межах всього 10 Гц навколо медіани 240.6 Гц) та екстремальну акустичну лабільність Юхима К., у якого зафіксовано різкий стрибок частоти основного тону до 322 Гц на третьому вимірюваному

сегменті. Такий може бути ознакою стану сильного психоемоційного та когнітивного напруження дитини. Акустична модель ASR, не адаптована до подібних раптових частотних девіацій дитячого голосу й інтерпретує такі ділянки як високочастотний шум, повністю спотворюючи результати орфографічного декодування. Саме тому описаний корпус, незважаючи на відносно невеликий обсяг, є цінним ресурсом для донавчання, адже дає системі зразки реального дитячого мовлення у всій його складності, включно з хезитаціями, суржиком, артикуляційними варіаціями і паралінгвістичними сигналами.

Висновки до третього розділу

Третій розділ присвячено практичній реалізації лінгвістичного анотування та аналізу характеристик сформованого корпусу. Підготовчий технічний етап передбачав декомпресію вхідних аудіофайлів у форматі .m4a до лінійного нестисненого формату PCM WAV через хмарну платформу Convertio. Верифікацію коректності конвертації здійснено шляхом контрольного відтворення та зіставлення тривалості файлів до і після перекодування, що повністю виключило ймовірність технічних розривів і часових зсувів.

Після імпорту до Praat кожен запис пройшов первинний візуальний та слуховий моніторинг у вікні «Sound + Spectrogram» для загальної орієнтації у структурі інтерв'ю та оцінки якості матеріалу. Для кожного запису створено чотирирівневий TextGrid. Мікросегментація здійснювалась вручну в інтерактивному редакторі «Sound + TextGrid Editor» за поєднаним акустично-спектральним критерієм. Як прагматичний поріг паузи встановлено 0,5 секунди: відрізки коротші за цю межу розглядаються як внутрішньофразові та позначаються маркером *пауз*, тоді як паузи понад 0,5 с визначаються як сигнал завершення комунікативної репліки і вимагають виставлення нової часової межі. Текстова фіксація здійснювалась відповідно до Правопису 2019 р., при цьому сегменти з незадовільною розбірливістю позначалися тегом *нерозб*. Загальні часові витрати на анотування одного запису становили 11–15 астрономічних годин.

Аналіз специфічних явищ дитячого мовлення виявив чотири основні категорії, що вимагали особливих анотаційних рішень: хезитаційні явища та паузації (найчастотніша категорія), повтори і самокорекції, перебивання та емоційні вигуки, паралінгвістичні прояви (сміх, хвилювання, усвідомлення). Розроблені анотаційні рішення для кожної категорії забезпечують збереження автентичного характеру дитячої мовленнєвої продукції для лінгвістичного аналізу та постачання ASR-системі верифікованих текстових даних.

Кількісний аналіз сформованого корпусу виявив такі характеристики: загальна тривалість — 76 хвилин 49 секунд; лексичний обсяг реплік дітей — 1 534 слова (токени); кількість названих ключових слів — 203. Найвищу мовленнєву продуктивність продемонструвала Віка К. (566 слів за 26 хв 33 с), найвищу дискурсивну щільність — Кіра Н. (511 слів за 22 хв 12 с), найнижчу продуктивність — Юхим К. (457 слів за 28 хв 04 с). Найчастотнішим маркером корпусу є *шт* (411 фіксацій), що відображає польовий характер збирання матеріалу, а розмічені сегменти з фоновим шумом є ресурсом для адаптаційного тренування моделі в реальних акустичних умовах.

Акустико-фонетичний аналіз засвідчив такі середні значення частоти основного тону: Віка К. — 297,6 Гц, Кіра Н. — 240,6 Гц, Юхим К. — 277,6 Гц, а середньозважений показник по корпусу — 271,9 Гц. Ці значення суттєво перевищують нормативні діапазони і чоловічого (100–150 Гц), і жіночого (180–220 Гц) дорослого голосу, що підтверджує принципову несумісність стандартних ASR-моделей із дитячим мовленням і обґрунтовує необхідність спеціалізованого адаптаційного тренування. Якісний аналіз соціолінгвістичних маркерів виявив контрастивні стратегії трьох мовців: Кіра Н. демонструє активну суржикову інтерференцію (36 маркерів *сур*) при повній відсутності суцільних російських сегментів; Юхим К. — тотальне домінування російськомовних сегментів (49 маркерів *р*) і стратегію комунікативної пасивності, що дзеркально підвищує інтенсивність реплік модератора до 45 фіксацій тегу *мод*. Ці розбіжності відображають реальний мовний ландшафт київських шкіл і є важливими для донавчання ASR.

ВИСНОВКИ

Виконана кваліфікаційна робота є комплексним теоретично-прикладним дослідженням, спрямованим на розробку й лінгвістичне анотування спеціалізованого корпусу усного дитячого мовлення в програмному середовищі Praat для потреб оптимізації систем автоматичного розпізнавання мовлення (ASR). Проведене дослідження дозволило зробити такі висновки.

1. На підставі аналізу праць вітчизняних і зарубіжних дослідників систематизовано поняттєво-категоріальний апарат корпусної лінгвістики. Встановлено, що повноцінний лінгвістичний корпус є системно зорганізованим, репрезентативним і обов'язково анотованим масивом даних. Принципова відмінність усних корпусів від письмових полягає у безперервності звукового потоку, що вимагає часової сегментації, фіксування просодичних і акустичних характеристик та обов'язкової транскрипції. Міжнародний досвід систем CHILDES, HomeBank і Spokes підтвердив, що аналітичний потенціал корпусу визначається якістю лінгвістичної розмітки, а не лінійним обсягом сирих даних.

2. Охарактеризовано специфіку дитячого мовлення як об'єкта лінгвістичного опису та автоматичного розпізнавання. Молодший школяр 7–8 років є специфічним суб'єктом мовленнєвої діяльності з принципово відмінним від дорослих акустичним профілем, у нього значно вища частота основного тону (220–300 Гц і вище), зміщені вгору значення формантних частот, нестабільна артикуляція, часті хезитації, незавершені синтаксичні конструкції та явища інтерференції. Разом ці чинники зумовлюють стабільно вищий показник Word Error Rate (WER) при обробці дитячого мовлення сучасними ASR-системами.

3. Обґрунтовано доцільність і принципи розробки спеціалізованого (закритого) корпусу для потреб ГО «Спільномова». Встановлено, що «золотий датасет», зосереджений на обмеженій лексиці конкретної методики та конкретній вікові й регіональній групі, є ефективнішим тренувальним матеріалом для дотреновування базових ASR-моделей (Whisper, wav2vec 2.0), ніж великі загальнонаціональні мовленнєві бази. Закритий статус корпусу обґрунтовано як

свідоме правове й етичне рішення відповідно до вимог Закону України «Про захист персональних даних» щодо біометричних даних неповнолітніх.

4. Охарактеризовано емпіричну базу дослідження та методику збирання матеріалу. З 67 первинних записів, зібраних у Ліцеї № 107 «Введенський» м. Києва у квітні–травні 2024 р., до фінальної вибірки відібрано три записи учнів 2Б класу (Віки К., Кіри Н. та Юхима К.), що утворюють репрезентативну контрастивну тріаду за мовним профілем (переважно україномовний, активно білінгвальний та фактично повністю російськомовний). Методика інтерв'ювання, розроблена ГО «Спільномова» з опорою на стандарти PPVT-5, EVT-3 і MAIN, забезпечила різноманітність мовлення.

5. Розроблено й апробовано систему транскрибування та лінгвістичного анотування дитячого мовлення. Система включає чотирирівневу TextGrid-архітектуру («Модератор», «Дитина», «Ключові слова (М)» та «Ключові слова (Д)»), нормативну основу (Правопис 2019 р.) та розгалужену систему маркерів п'яти функціональних груп: мовних відхилень (*р*, *сур*), способу вимови (*роз*, *пскл*, *шеп*, *гарк*, *спот*, *дслов*, *затр*, *об*), хезитаційних явищ (*е*, *мммм*, *звроз*, *пауз*, *м-м*, *вд*), паралінгвістичних проявів (*см*, *сміх*, *усм*, *хм*, *хв*) та фонового шуму (*шт*, *шкор*, *скр*, *стук*, *шар*, *крикд*). Запропонована система є відтворюваною та придатною для застосування при розширенні цього чи проєктуванні аналогічних корпусів.

6. Реалізовано практичну процедуру лінгвістичного анотування трьох аудіозаписів у Praat. Технічна підготовка матеріалу передбачала конвертацію файлів .m4a до формату PCM WAV. Мікросегментація здійснювалась за акустично-спектральним критерієм з порогом паузи 0,5 с; текстова фіксація виконувалась згідно з Правописом 2019 р. Загальні часові витрати на анотування одного запису становили 11–15 годин, що відображає трудомісткість корпусних досліджень спонтанного мовлення.

7. Проведено кількісний і якісний аналіз сформованого корпусу. Загальна тривалість — 76 хвилин 49 секунд, лексичний обсяг дитячих реплік — 1 534 слова, 203 названих ключових слова. Середній показник частоти основного тону

по корпусу — 271,9 Гц, що суттєво перевищує нормативні діапазони дорослих мовців і підтверджує принципову несумісність стандартних ASR-моделей із дитячим мовленням. Найчастотнішим маркером є *шт* (411 фіксацій), присутність якого у розмічених сегментах є перевагою для адаптаційного тренування щодо реальних акустичних умов. Якісний аналіз соціолінгвістичних маркерів виявив суттєві відмінності між мовними профілями трьох опитаних, що відображають мовний ландшафт київських шкіл.

Практичне значення отриманих результатів полягає у безпосередньому формуванні бази лінгвістичних даних, а саме розмічених аудіофайлів у форматі .wav та відповідних TextGrid-об'єктів, готових до інтеграції в процеси дотреновування ASR-моделей для автоматичного розпізнавання українського дитячого мовлення. Розроблені інструкції та принципи лінгвістичного анотування забезпечують відтворюваність методики при подальшому розширенні корпусу або проєктуванні аналогічних датасетів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бацевич Ф. С. Основи комунікативної лінгвістики: Підручник. -К.: Видавничий центр «Академія», 2004. - 344 с. – ISBN 966-580-172-4
2. Бобкова Т. В. Корпус текстів: основні аспекти визначення. Науковий вісник кафедри Юнеско КНЛУ, 2014. Випуск 29. С. 11–20. URL: https://www.mova.info/corpus_papers/bobkova-corpus.pdf
3. Гертлер К. Зоопарк! Віммельбух. Харків : ARTBOOKS, 2016. ISBN 978-617-7395-02-6. URL: https://artbooks.ua/zoopark-vimmel-buh
4. ГО «Спільномова» : офіційний сайт. URL: https://www.spilnomova.org/
5. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник //К.: Видавничо-поліграфічний центр «Київський університет. – 2008. – Т. 351.
6. Демська-Кульчицька О. Дещо про класифікацію текстових корпусів. Наукові записки. Серія: Мовознавство. 2004. 1(11). С. 153–157. URL: https://ekmair.ukma.edu.ua/server/api/core/bitstreams/ae0b2474-1b67-4c03-b6b4-66707e5e94a0/content
7. Демська-Кульчицька. О. Базові поняття корпусної лінгвістики // Українська мова, 2003, 1 (6): 40–45. URL: https://ekmair.ukma.edu.ua/server/api/core/bitstreams/d30cf29c-acbf-4221-bd40-03b04f5f9bd5/content
8. Жуковська В. В. Вступ до корпусної лінгвістики. Житомир: ЖДУ ім. І. Франка, 2013. 142 с. URL: https://eprints.zu.edu.ua/18909/1/korpusna_lingv.pdf#page=58.09
9. Захист персональних даних неповнолітніх. URL: https://legalaid.wiki/index.php/Захист_персональних_даних_неповнолітніх
10. Іщенко О.С. Вплив іменників і дієслів на паузування в сному мовленні. Українська мова. 2020. No 2 (74). С. 45—58. URL: https://iul-nasu.org.ua/pdf/ukrmova/2_20/6.pdf
11. Корпусна лінгвістика / Є. А. Карпіловська // Енциклопедія Сучасної України [Електронний ресурс] / редкол. : І. М. Дзюба, А. І. Жуковський, М.

Г. Железняк [та ін.] ; НАН України, НТШ. – Київ: Інститут енциклопедичних досліджень НАН України, 2014. – Режим доступу: https://esu.com.ua/article-5251.

12. Корпусна лінгвістика [Текст] / В. А. Широков [та ін.] ; відп. ред. В. А. Широков ; НАН України, Укр. мовно-інформ. фонд. - К. : Довіра, 2005. – 472 с.: табл. - Бібліогр.: с. 459–467. – ISBN 966-507-189-0
13. Про захист персональних даних : Закон України від 01.06.2010 р. № 2297-VI. URL: https://zakon.rada.gov.ua/laws/show/2297-17#Text
14. Український правопис. Київ, 2019. URL: https://mon.gov.ua/static-objects/mon/sites/1/zagalna%20serednya/Pravopys.2019/ukr.pravopys-2019.pdf.
15. Assessment of Narrative Abilities in Bilingual Children*Natalia Gagarina, Daleen Klop, Sari Kunnari, Koula Tantele, Taina Välimaa, IngridaBalčiūnienė, Ute Bohnacker, Joel WaltersGagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. &Walters, J. 2015. Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J.de Jong & N. Meir (eds.) Methods for assessing multilingual children: Disentangling bilingualism from language impairment. Bristol, UK: Multilingual Matters.1 Introduction and overview. URL: https://www.researchgate.net/publication/348461736_MAIN_Multilingual_Assessment_Instrument_for_Narratives_-_Revised
16. Average Speaking Frequencies: F0 Norms by Age, Sex, and Hormonal Status. URL: https://www.voicescience.org/lexicon/average-speaking-frequencies/#3-children-and-pre-pubescent-voices
17. Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // NeurIPS. 2020. URL: https://proceedings.neurips.cc/paper/2020/file/92d1e1eb1cd6f9fba3227870bb6d7f07-Paper.pdf
18. Bhardwaj, Vivek & Othman, Mohamed & Kukreja, Vinay & Belkhier, Youcef & Bajaj, Mohit & Goud, B. & Rehman, Ateeq & Shafiq, Muhammad & Hamam,

Habib. (2022). Automatic Speech Recognition (ASR) Systems for Children: A Systematic Literature Review. Applied Sciences. 12. 10.3390/app12094419. URL: https://www.researchgate.net/publication/360131327_Automatic_Speech_Recognition_ASR_Systems_for_Children_A_Systematic_Literature_Review

19. Boersma, Paul & Weenink, David. (2007). PRAAT: Doing phonetics by computer (Version 5.3.51). URL: https://www.researchgate.net/publication/259810776_PRAAT_Doing_phonetics_by_computer_Version_5351
20. Bruner J. Child's Talk: Learning to Use Language. New York: W. W. Norton & Company, 1983. 144 p. URL: https://ndl.ethernet.edu.et/bitstream/123456789/16634/1/pdf58.pdf
    – DOI: 10.5281/zenodo.19997543
21. Dunn D. M. Peabody Picture Vocabulary Test | Fifth Edition (PPVT-5). 2018. URL: https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Academic-Learning/Peabody-Picture-Vocabulary-Test-%7C-Fifth-Edition/p/100001984
22. Eigsti I.-M. Peabody Picture Vocabulary Test. Springer Science+Business Media LLC 2017 F.R. Volkmar (ed.), Encyclopedia of Autism Spectrum Disorders, DOI 10.1007/978-1-4614-6435-8_531-3. URL: https://www.researchgate.net/profile/Inge-Marie-Eigsti/publication/318874497_Peabody_Picture_Vocabulary_Test/links/5bcf9750299bf1a43d9c6165/Peabody-Picture-Vocabulary-Test.pdf.
23. Gerosa, Matteo & Giuliani, Diego & Narayanan, Shrikanth & Potamianos, Alexandros. (2009). A review of ASR technologies for children's speech. Proceedings of the 2nd Workshop on Child, Computer and Interaction, WOCCI '09. 10.1145/1640377.1640384. URL: https://www.researchgate.net/publication/228625959_A_review_of_ASR_technologies_for_children's_speech

24. Gut, U. Praat Manual Linguistics Laboratory University of Augsburg. 2008. URL: https://www.uni-muenster.de/imperia/md/content/englischesseminar/gut/praat_manual1.pdf
25. Kennedy G. Introduction to Corpus Linguistics. - London-New-York: Longman, 1998. – 309 p. URL: https://www.scribd.com/document/507842286/An-Introduction-to-Corpus-Linguistics?v=0.389
26. Kovalyov A. Допис у соціальній мережі Facebook від 22 квітня 2024 р. URL: https://www.facebook.com/andriy.kovalyov.5/posts/pfbid02bSu6j2KA4Df4jRvi2L8nzs2sCTtWtSxzHjxu5fVFHZXJDtYnr6wEvA4RyhtETcs5l
27. Kovalyov A. Допис у соціальній мережі Facebook від 29 листопада 2023 р. URL: https://www.facebook.com/andriy.kovalyov.5/posts/pfbid0AxQRVP1r2aZcsfGmFei3LZvyMmrbxa38vDy6jJorSAV8QsKNfSCbuasvZCAKvvZml
28. Kovalyov A. Допис у соціальній мережі Facebook від 3 липня 2024 р. URL: https://www.facebook.com/andriy.kovalyov.5/posts/pfbid0n5ieGuxQvNnZqCf3oCmS3sEZXkSFGHh5h8p3pJ5nrErUwj79RWASpVLLA6p3R8Ucl
29. Lennes, M 2009, Segmental features in spontaneous and read-aloud Finnish. in V D S R U (ed.), Phonetics of Russian and Finnish : general description of phonetic systems : experimental studies on spontaneous and read-aloud speech. Peter Lang , Frankfurt am Main, pp. 145-166. URL: https://helda.helsinki.fi/server/api/core/bitstreams/0af48719-344c-40ed-80fb-5149ac793659/content#page=4.00
30. Lennes, M 2019, SpeCT 2.0 – Speech Corpus Toolkit for Praat. in K Simov & M Eskevich (eds), CLARIN Annual Conference 2019 : Proceedings. CLARIN, Leipzig, pp. 147-149, CLARIN Annual Conference, Leipzig, Germany, 30/09/2019. URL: https://helda.helsinki.fi/server/api/core/bitstreams/b47ebb96-32d9-4a96-bd42-d7e34373e801/content#page=3.33
31. M4A to WAV Converter. URL: https://convertio.co/m4a-wav/.
32. MacWhinney B. The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000. URL:

https://www.researchgate.net/publication/243532847_The_CHILDES_project_tools_for_analyzing_talk

33. MAIN (Multilingual Assessment Instrument for Narratives). URL: https://main.leibniz-zas.de/
34. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2011. 312 p. URL: https://nlp.csie.org/~sound/Corpus%20Linguistics%20Method%20Theory%20and%20Practice.pdf#page=9.00
35. Pęzik, Piotr. “Spokes – a Search and Exploration Service for Conversational Corpus Data.” In Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands, 99–109. Linköping Electronic Conference Proceedings. Linköping University Electronic Press, Linköpings universitet, 2015. URL: https://ep.liu.se/ecp/116/009/ecp15116009.pdf
36. Radford A., Kim J. W., Xu T., Brockman G. et al. Robust Speech Recognition via Large-Scale Weak Supervision [OpenAI Whisper] // arXiv:2212.04356. 2022. URL: https://proceedings.mlr.press/v202/radford23a/radford23a.pdf
37. Simmering P. How to Get Gold Standard Data for NLP. 10.03.2024. URL: https://simmering.dev/blog/gold-data/.
38. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991. 179 p. URL: https://www.scribd.com/document/512884816/Corpus-Concordance-Collocation-by-John-Sinclair-Z-lib-org
39. Sinclair, J. 2005. "Corpus and Text - Basic Principles" in Developing Linguistic Corpora: a Guide to Good Practice, ed. M. Wynne. Oxford: Oxbow Books: 1–16. DOI https://doi.org/20.500.14106/dlc
40. Slobin, D.I. (Ed.). (1985). The Crosslinguistic Study of Language Acquisition: Volume 1: the Data (1st ed.). Psychology Press. https://doi.org/10.4324/9781315802541
41. Spokes. URL: https://spokes.clarin-pl.eu/home

42. STEC Guidance on Expressive Vocabulary Test, 3rd Edition (EVT-3). 2018. URL: https://www.sasc.org.uk/media/yjkd0rnc/expressive-vocabulary-test-3-guidance-sasc-may-2021.pdf
43. STEC Guidance. PPVT-5 (Peabody Picture Vocabulary Test, Fifth Edition). November 2025. URL: https://www.sasc.org.uk/media/qnxnr3eh/ppvt5-guidance-november-2025.doc.
44. Tomasello M. (2000). The item-based nature of children's early syntactic development. Trends in Cognitive Sciences, 4(4), 156–163. URL: https://glowlinguistics.org/37/pdf/tomasello2000.pdf
45. VanDam M., Warlaumont A. S., Bergelson E., Cristia A., Soderstrom M., De Palma P., MacWhinney B. HomeBank: An Online Repository of Daylong Child-Centered Audio Recordings. Seminars in Speech and Language. 2016. Vol. 37, No. 2. P. 128–142. URL: https://www.thieme-connect.de/products/ejournals/html/10.1055/s-0036-1580745#jumpto-the-problem-of-distributed-and-sequestered-data
46. Williams K. T. Expressive Vocabulary Test | Third Edition (EVT-3). 2018. URL: https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Academic-Learning/Expressive-Vocabulary-Test-%7C-Third-Edition/p/100001982

ДОДАТКИ

Додаток 1. Система символів (маркерів лінгвістичного анотування усного дитячого мовлення) для розмітки спеціалізованого корпусу

ПОЗНАЧЕННЯ СПОСОБУ ВИМОВИ

Позначення стосується тільки одного наступного слова (ставиться перед ним):	
о	Обрив — слово обривається (обрізаний звуковий файл, диктора перебивають, диктор не договорює тощо). Фіксується лише та частина слова, яка реально звучить.
мод	Модероване мовлення — слово не договорюється навмисно. Використовується як підказка дитині, щоб вона самостійно продовжила слово, якщо знає його.
об	Обмовка — диктор помиляється й відразу сам себе виправляє. Також використовується у випадку просто неправильної вимови деяких звуків у слові. Фіксується лише та частина, яка реально звучить.
роз	Розтягування — слово вимовляється зі значним розтягуванням одного або кількох звуків. У текстове поле вноситься літературний варіант слова.
звроз	Розтягування одного звука, що відбувається безпосередньо в процесі думання / паузи роздумів.
пскл	По-складова вимова — слово вимовляється чітко за складами. Вноситься літературний варіант слова.
пбкв	Побуквенна вимова — слово вимовляється по буквах.
з	Заїкання. Вноситься нормативний літературний варіант слова.
шеп	Шепелявість при вимові. Вноситься літературний варіант слова.
затр	Затримка — слово вимовляється з невеликою внутрішньою затримкою (наприклад, *в-она*).

гарк	Гаркавість — неправильна або специфічна вимова звуків «р» і «л». Вноситься літературний варіант слова.
зклдн	Вимова слів при закладеному носі мовця. Вноситься літературний варіант слова.
клуб	Вимова слів із відчуттям «клубка в горлі». Вноситься літературний варіант слова.
деф	Дефект слова — не чутно закінчення слова через занадто швидку вимову, погану якість запису або одночасне говоріння обох мовців. Фіксується лише та частина, яку чітко чутно.
спот	Спотворене слово — дитина неправильно вимовляє слово через те, що воно є занадто важким для неї або містить звуки, які вона ще не вміє чітко артикулювати.
дслов	Використання дитиною неправильних, ненормативних граматичних або морфологічних форм слів.

ПОЗНАЧЕННЯ ФОНУ (ТЛА)

Позначення стосується цілого сегменту (рядка) і ставиться на його початку	
шт	Шум на тлі — будь-який сторонній акустичний шум, на фоні якого говорить диктор (наприклад, коли діти голосно кричать у коридорі).

ХЕЗИТАЦІЇ ТА СТИЛЬ МОВЛЕННЯ

Означає відповідні звуки в записі, які видає мовець (ставиться на місці звука)	
е	Екання, акання, икання, укання тощо (наповнена пауза).
у-у	Звуковий жест заперечення.
угу	Угукання, агакання, довге «-а-» (у значенні «ясно / зрозуміло»).
вд	Фізіологічний вдих або видих мовця.
мммм	Звук у значенні «задумався» під час пошуку лексеми.
мм	Коротке запитання (типу м?, а?).
сьорб	Звук сьорбання.

св	Звук свисту.
ковт	Звук ковтання.
м-м	Звук вагання мовця перед вибором слова.
ммм	Звуковий жест типу «понятно».
м-мм	Звуковий жест типу «ні».
ааа	Хезитаційний вокалізований звук «ааа».
зв	Будь-який дивний, нетиповий чи незрозумілий звук, який видав мовець.
ім	Навмисна імітація якогось звуку мовцем.
м	Короткий звук «м».
усм	Вимова слова чи речення з усмішкою, трошки сміючись. Може стосуватися як окремого слова, так і цілого сегменту.
син	Синхронне мовлення — одночасне говоріння обох мовців (стосується одного або кількох накладених слів).
пл	Звук плямкання.
хм	Хмикання.
см	Одиночний або короткий сміх.
смд	Сміх із вираженим вдихом чи видихом (із диханням).
тих	Мовець говорить дуже тихо (локальна позначка).
йой	Йойкання.
виг	Спонтанний вигук.
цок	Цокання ротом / язиком.
чхи	Чхання.

ПАРАЛІНГВІСТИЧНІ ЯВИЩА

Позначення стосується цілого сегменту	
спів	Спів (диктор проспівує репліку або її частину).
шепіт	Шепіт (диктор повністю переходить на шепотіння).

сміх	Виражений сміх або виражена усміхнена вимова протягом усього сегменту.
усвід	Інтонація усвідомлення (наприклад: «Аааа, так он воно як!» після пояснення модератора).
розч	Інтонація розчарування, розпачу (коли мовець уже ледь не плаче або перебуває у стані напівплачу, наприклад: «Ну блин, я не слышу тебя»).
сарк	Саркастична інтонація мовця.
хв	Виражене, чітко помітне хвилювання в голосі.
зл	Злість у голосі, експресія агресії або явна роздратованість диктора.
тих	Мовець тихо говорить протягом усього інтервалу рядка.

ШУМИ ТА НЕРОБОЧІ ВІДРІЗКИ

Позначення стосується цілого сегменту або конкретного моменту шуму	
шум	Невизначений, акустично розмитий сторонній шум.
шкор	Шум, що долинає зі шкільного коридору.
стук	Чіткий стукіт (по парті, у двері тощо).
скр	Скрип (шкільного стільця, дверей тощо).
пауз	Фізична незаповнена пауза під час говоріння всередині одного сегменту.
зс	Специфічний звук перетягування стільця по підлозі.
дих	Голосне дихання на фоні запису (коли дитина дихає безпосередньо в мікрофон). Якщо фіксується у репліках модератора — це зазвичай теж дихання дитини, що потрапило в канал.
шм	Шмигання носом.
гов	Стороннє говоріння на тлі запису.
крик	Невизначений голосний крик.
крикд	Голосний крик сторонньої дитини (наприклад, крики дітей у коридорі під час перерви).

дит	Звуки, видані дитиною, що зафіксовані під час репліки модератора.
шар	Шарудіння папером, картками чи сторінками книги.
мд	Звуки, короткі вигуки або реакції перебивання з боку модератора всередині репліки дитини.
нерозб	Нерозбірливо — диктор говорить абсолютно нерозбірливо (маркер вноситься в поле інтервалу замість текстової стенограми).
приск	Прискорення — аномальне прискорення темпу мовлення диктора (вноситься замість слів диктора у сегменті).

Додаток 2. Електронні матеріали спеціалізованого корпусу дитячого мовлення

Папка містить 3 аудіофайли у форматі .wav та 3 синхронізовані файли розмітки TextGrid із чотирма інтервальними рівнями анотації.

https://drive.google.com/drive/folders/1x73V33EIhFHDH7pbPHskND8SCZWtQykd?usp=drive_link

Додаток 3. Результати анкетування батьків

Папка містить 1 файл у форматі .xlsx із результатами опитування, проведеного серед батьків. Усю конфіденційну інформацію та персональні дані було видалено з міркувань безпеки.

https://drive.google.com/drive/folders/1DvorpJ5E4Z9TpD-3VeyS8rw-US8n4Khn?usp=drive_link