

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

**УКЛАДАННЯ ЕЛЕКТРОННОГО ПЕРЕКЛАДАЦЬКОГО
ЯПОНСЬКО-УКРАЇНСЬКОГО СЛОВНИКА
ОНОМАТОПЕЇЧНОЇ ЛЕКСИКИ**

Кваліфікаційна робота бакалавра
студента VI курсу
ОПП «Прикладна (комп'ютерна)
лінгвістика та англійська мова»
спеціальності 035 «Філологія»
спеціалізації 035.10 «Прикладна лінгвістика»
галузі знань 03 «Гуманітарні науки»

Олексія ЛАРІОНОВА

Науковий керівник:

доктор філологічних наук, професор кафедри
української мови та прикладної лінгвістики

Наталія ДАРЧУК

«Допущено до захисту»

Протокол засідання кафедри

української мови та прикладної лінгвістики

від 28.05.2025 № 15

Завідувач кафедри _____ **Сергій РІЗНИК**

КИЇВ — 2025

АНОТАЦІЯ

Ларіонов О. Укладання електронного перекладацького японсько-українського словника ониматопеїчної лексики. — Кваліфікаційна робота освітнього рівня «Бакалавр» на здобуття освітнього-кваліфікаційного рівня вищої освіти спеціальності 035 «Філологія» спеціалізації «Прикладна лінгвістика». — Київський національний університет імені Тараса Шевченка. Київ, 2025.

Робота присвячена створення двомовного (японсько-українського) словника для ілюстрації ониматопеїчного класу лексики обох мов та здійснення пошуку міжмовних еквівалентів. У роботі визначено термінологію, якою послуговуються для опису ониматопеїчних (звуконаслідувальних) одиниць, окреслено лексико-семантичні відмінності зазначених одиниць від вигуків, проаналізовано класифікації звуконаслідувань.

На основі даних з 6 джерел («Словник української мови в 20 т.», «Словник української мови за ред. Б. Грінченка», «Великий тлумачний словник сучасної української мови вид. “Перун”», «Великий електронний словник української мови», блог «livejournal», доробок видавництва «Nasha Idea») створено базу даних української звуконаслідувальної та вигуків лексики, на основі 3 джерел («Японсько-український тематичний словник ониматопеїчної лексики», «nihongoresources», «Japanese Manga SFX») — аналогічну базу даних японських звуконаслідувань та вигуків. Використано засоби мови програмування Python для автоматизованої обробки матеріалів дослідження. Побудовано функціонал пошуку еквівалентів між японськими та українськими ониматопами шляхом векторизації словникових визначень та обрахування косинусної семантичної подібності, створено користувацький інтерфейс електронного словника та подано його у вигляді вебсайту.

Ключові слова: ониматопея, звуконаслідування, вигук, електронний словник, перекладацький словник, гісей-го, гіон-го, гітай-го.

ABSTRACT

Larionov O. Compiling a Digital Japanese-Ukrainian Bilingual Dictionary of Onomatopoeic Words. – Qualification work for the bachelor’s degree in higher education in specialty 035 «Philology», specialization «Applied Linguistics». – Taras Shevchenko National University of Kyiv. Kyiv, 2025.

The paper is devoted to compiling a bilingual (Japanese-Ukrainian) dictionary aimed at illustrating the onomatopoeic lexic of both languages and identifying cross-lingual equivalents. Specific terminology used in regards to onomatopoeic (sound-imitating) units by Ukrainian and Japanese linguists as well as semantic differences between onomatopoeia and interjections and their classifications are examined.

Based on the material from 6 sources (“Dictionary of the Ukrainian Language in 20 Volumes”, “Dictionary of the Ukrainian Language edited by B. Hrinchenko”, “Comprehensive Explanatory Dictionary of Modern Ukrainian by Perun Publishers”, “Comprehensive Electronic Dictionary of Ukrainian”, the LiveJournal blog, and materials by Nasha Idea publishing house) the database of Ukrainian onomatopoeia and interjections is compiled. Similarly, based on the material from 3 sources (“Japanese-Ukrainian thematic dictionary of onomatopoeic words”, “nihongoresources”, “Japanese Manga SFX”) a database is compiled for Japanese onomatopoeia and interjections. Programming language Python is used for the automatized processing of aforementioned sources. Search functionality for identification of cross-lingual equivalents is constructed by means of word-embedding and cosine semantic similarity evaluation. The user interface of the digital dictionary is created and presented in the form of a website.

Keywords: onomatopoeia, sound imitation, interjection, digital dictionary, bilingual dictionary, gisei-go, gion-go, gitai-go.

ЗМІСТ

ВСТУП.....	6
РОЗДІЛ 1. СПЕЦИФІКА ПІДХОДІВ ДО ВИВЧЕННЯ ОНОМАТОПІВ В УКРАЇНСЬКОМУ МОВОЗНАВСТВІ.....	10
1.1. Різні аспекти дослідження звуконаслідувань.....	10
1.2. Метамова дослідження.....	11
1.3. Відмінність звуконаслідувальних слів та вигуків в українській мові.....	12
1.4. Підходи до класифікації звуконаслідувальних слів в українській мові...	14
Висновки до Розділу 1.....	20
РОЗДІЛ 2. СТАНОВИЩЕ ОНОМАТОПЕЇЧНОЇ ЛЕКСИКИ В ДОСЛІДЖЕННЯХ З ЯПОНІСТИКИ.....	21
2.1. Дослідження японських ономатопів українськими та японськими науковцями.....	21
2.2. Звукосимволічність японської мовної традиції.....	22
2.3. Структурно-формальна класифікація японських ономатопів.....	25
2.4. Семантичні групи японських ономатопів.....	27
2.5. Проблеми японсько-українського перекладу.....	29
Висновки до Розділу 2.....	31
РОЗДІЛ 3. СТВОРЕННЯ БАЗИ ДАНИХ УКРАЇНСЬКИХ ОНОМАТОПІВ	32
3.1. Українська ономатопеїчна лексика: експертна оцінка електронних лексикографічних джерел.....	32
3.2. Екстракція української ономатопеїчної лексики із електронних ресурсів та редагування бази даних.....	34
3.2.1. Словник української мови в 20 томах.....	34
3.2.2. Словник української мови за ред. Б. Грінченка.....	37
3.2.3. Великий тлумачний словник сучасної української мови вид. «Перун».....	40

3.2.4. Великий електронний словник української мови.....	42
3.2.5. Інші електронні джерела.....	43
3.3. Лінгвістичний аналіз бази даних української ономатопеїчної лексики...	45
Висновки до Розділу 3.....	48
РОЗДІЛ 4. СТВОРЕННЯ БАЗИ ДАНИХ ЯПОНСЬКИХ ОНОМАТОПІВ...	49
4.1. Японська ономатопеїчна лексика: експертна оцінка електронних лексикографічних джерел.....	49
4.2. Екстракція японської ономатопеїчної лексики із електронних ресурсів та редагування бази даних.....	50
4.2.1. Японсько-український тематичний словник ономатопеїчної лексики.....	50
4.2.2. База даних ресурсу «nihongoresources».....	53
4.2.3. Користувацький електронний словник «Japanese Manga SFX».....	54
4.2.4. Інтеграція транскрипції ромаджі до словникових статей.....	58
4.3. Лінгвістичний аналіз бази даних японської ономатопеїчної лексики.....	59
Висновки до Розділу 4.....	61
РОЗДІЛ 5. ПОБУДОВА СИСТЕМИ ПОШУКУ УКРАЇНСЬКИХ ВІДПОВІДНИКІВ ДЛЯ ЯПОНСЬКИХ ОНОМАТОПІВ.....	62
5.1. Обрахування семантичної подібності.....	62
5.2. Побудова функцій пошуку в базі даних.....	64
5.3. Побудова користувацького інтерфейсу електронного словника.....	68
5.4. Демонстрація результатів.....	72
Висновки до Розділу 5.....	75
ВИСНОВКИ.....	76
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	78
ДОДАТКИ.....	84

ВСТУП

Ономатопейчна лексика наявна в будь-якій мові світу. Як зазначає дослідниця О. Валігура: «Ономатопи (звуконаслідувальні слова) є одиницями звуконаслідувальної підсистеми звукозображувальної системи мови, в основі яких лежить фонетично мотивований зв'язок між фонематичною формою слова та звуковою ознакою денотата» [3, с. 32]. Ця лексика є відомою та інтуїтивно впізнаваною для кожного носія певної мови, має широке застосування в повсякденному мовленні і характерна для художнього стилю, особливо для поетичних творів завдяки своїй лаконічності у поєднанні з глибоким емоційним підтекстом, який може бути нею висловлений.

В українському мовознавстві немає єдиної загальноприйнятої класифікації ономатопів, а відтак укладених академічних лексикографічних видань, які б ілюстрували все різнобарв'я звуконаслідувальної лексики, яка була частиною лексико-семантичного арсеналу з давніх часів і функціонує в українській мові сьогодні.

Особливо важливою й актуальною постає проблема ономатопів у зв'язку з перекладом на інші мови, зокрема японську. Японська мова має надзвичайно широку мовленнєву традицію у контексті звукового символізму, її носії послуговуються значно більшою кількістю можливостей, які надає ономатопейчний пласт лексики, зокрема в художніх творах. Це покриття широкого кола семантичних понять ономатопейчною лексикою, включаючи передавання емоцій, фізичних станів людини, зовнішніх характеристик тощо, які насправді не є звуками у своєму реальному прояві [12].

У зв'язку з цим **актуальність** нашої роботи зумовлена, по-перше, відсутністю укладених словникових видань та електронних ресурсів з повним реєстром звуконаслідувальних слів в українській мові і джерельною ілюстрацією, і по-друге, відсутністю джерел, в яких наводилися б міжмовні еквіваленти для японської та української мов.

Мета кваліфікаційної роботи: створити електронний ресурс, у якому буде подаватися паралельно список українських та японських ономатопів з

позначенням дефініцій та семантичних класів з пошуковим функціоналом. Щоб досягти цієї мети, необхідно виконати такі **завдання**:

- 1) Дослідити особливості функціонування ониматопеїчної лексики української мови на основі наявних наукових досліджень;
- 2) Визначити підходи до класифікації структурних і семантичних класів ониматопів;
- 3) Ознайомитися з науковою літературою, присвяченою проблемі ониматопів в японській мові;
- 4) Визначити труднощі перекладу зазначеної лексики з японської мови на українську;
- 5) Укласти лексикони ониматопів української та японської мови в межах електронної реляційної бази даних;
- 6) Здійснити лінгвістичний аналіз бази даних українсько-японських ониматопів;
- 7) Розробити інтернет-ресурс з функціоналом електронного словника.

Об'єктом дослідження є українські та японські ониматопи (звуконаслідувальні слова).

Предметом дослідження є перекладацька лексикографічна модель української та японської ониматопеїчної лексики та структура відповідного перекладацького електронного словника.

Методи дослідження: метод семантичної класифікації; метод комп'ютерного моделювання; метод об'єктно-орієнтованого програмування.

Інформаційна база дослідження: мови програмування Python [66] та JavaScript [49], програмні бібліотеки sqlite3 [73], re [68], BeautifulSoup [42], aiohttp [50], asyncio [41], csv [46], itertools [56], dataclasses [47], typing [75], collections [43], json [58], fugashi [52], pykakasi [67], Flask [51], pandas [65], numpy [63], SentenceTransformer [70], мова розмітки вебсайтів HTML [53] у поєднанні з мовою стилів CSS [48].

Практичне значення цієї роботи полягає в тому, що електронна система створеного перекладацького словника може стати ефективним і зручним

інструментом для правильного добору еквівалентів ономотопів у роботі перекладачів з японської на українську. Також така розробка має практичну цінність для всіх, хто зацікавиться особливостями функціонування української ономотопеїчної лексики, вивченням її коректного вживання, а також для дослідження її фоносемантичних характеристик.

Матеріал дослідження: «Словник української мови в 20 т.» [38], «Словник української мови за ред. Б. Грінченка» [36], «Великий тлумачний словник сучасної української мови вид. “Перун”» [32], «Великий електронний словник української мови» [33], блог «livejournal» [59], список ономотопеїчної лексики з доробку видавництва «Nasha Idea» [60]. «Японсько-український тематичний словник ономотопеїчної лексики» [37], «Словник української мови за ред. Б. Грінченка» [36], база даних ресурсу «nihongoresources» [61], користувацький електронний словник «Japanese Manga SFX» [57].

Структура кваліфікаційної роботи:

Кваліфікаційна робота бакалавра складається з п'яти розділів, списку використаних джерел та додатків.

У першому розділі проаналізовано низку теоретичних досліджень українських мовознавців. Визначено термінологію, якою послуговуються для опису ономотопеїчних (звуконаслідувальних) одиниць, окреслено лексико-семантичні відмінності зазначених одиниць від вигуків, проаналізовано класифікації звуконаслідувань за різними принципами.

У другому розділі проведено огляд досліджень ономотопеїчної лексики японської мови на основі праць українських та японських лінгвістів, визначено аспекти розгляду ономотопеїчної лексики японської мови в лінгвістичних дослідженнях, роль японських ономотопів у фоносемантиці японської мови, класифікацію японських ономотопів за формально-структурним та за семантичним показником, проаналізовано їхню семну структуру.

У третьому розділі проаналізовано вміст шести джерел для добору матеріалів поповнення бази даних української ономотопеїчної лексики,

використано засоби мови програмування Python [66] для автоматизованої обробки вмісту вказаних джерел та їхнього внесення до реляційної бази даних.

У четвертому розділі проаналізовано вміст трьох джерел для добору матеріалів поповнення бази даних японської ономотопеїчної лексики, використано засоби мови програмування Python [66] для автоматизованої обробки вмісту вказаних джерел та їхнього внесення до реляційної бази даних.

У п'ятому розділі розглянуто практичні аспекти побудови системи електронного словника за наявності попередньо створеної бази даних, вирішено питання побудови функціонали пошуку еквівалентів між японськими та українськими ономотопами шляхом векторизації словникових визначень та обрахування косинусної семантичної подібності, побудовано користувацький інтерфейс електронного словника, що реалізує здійснення пошуку японських ономотопів та виведення українських відповідників та пошук українських ономотопів з бази даних.

Під час роботи було опрацьовано 75 джерел, з яких 33 українською мовою, 32 англійською мовою, 10 японською мовою.

РОЗДІЛ 1. СПЕЦИФІКА ПІДХОДІВ ДО ВИВЧЕННЯ ОНОМАТОПІВ В УКРАЇНСЬКОМУ МОВОЗНАВСТВІ

1.1. Різні аспекти дослідження звуконаслідувань

Звуконаслідувальна лексика завжди була у фокусі уваги лінгвістів, причому в межах різних мовознавчих дисциплін. Про це зазначають І. Гаценко [4], Л. Богдан [1], Л. Мацько [17].

Насамперед звуконаслідувальна лексика цікава як специфічний лексичний шар, який має також свої особливості, співвідносні з різними функціями в реченні. Принагідно зазначимо, що вигук не дорівнює звуконаслідувальному слову, що показують деякі дослідження, зазначені нижче [8].

Нас зацікавила ще й проблема мотивованості слова, відома ще з античних часів, яка була в полі зору таких видатних мовознавців, як В. фон Гумбольдт та О.Потебня. О.Потебня зазначав, що існує «зовнішня» та «внутрішня» форми слова, між якими наявний нерозривний зв'язок, який, проте, з часом втрачався з різних причин. Внутрішня форма слова дозволяє зрозуміти, чому значенню відповідає той чи інший звуковий образ, що також корелює з фактом різних мотиваційних ознак слова в різних мовах у зв'язку з відмінними особливостями фонетичних систем [23]. Виходячи з цієї теорії, звуконаслідування є словами, в яких мотивованість прослідковується чи не найбільш явно та прозоро.

Якщо ж апелювати до досліджень Ф. де Соссюра, то він визначав тему звуконаслідувальних слів як «периферійну» та висував тезу про немотивованість мовного знака, аргументуючи це тим, що звуконаслідування не є «органічними елементами мови та займають кількісно незначне місце у її системі». Однак, з іншого боку, сам Ф. де Соссюр визнавав, що звуконаслідувальні слова змінюються під впливом мовної еволюції та її законів, набувають рис, характерних для решти мовних знаків, тим самим по суті визнаючи їхню повноправність [8, с. 130].

Не можна оминути також важливості звуконаслідувальної лексики у контексті фоносемантики. В цій мовознавчій дисципліні вивчається

звукотобразувальний потенціал мови, встановлюються залежності між фонетичною формою слова та денотатом, його ознаками, в тому числі психоемоційними [21]. Звуконаслідування є ключовим поняттям фоносемантики. Засновником цієї науки вважається С. Воронін, який у праці «Основи фоносемантики» виклав її основні засади [20].

О. Кобелянська, досліджуючи звуковий символізм (фоносемантику) японської ониматопейчної лексики, зазначає: «У процесі численних лінгвостатистичних експериментів дослідники довели, що для переважної більшості природних носіїв української мови відповідні приголосні звуки цих мов асоціюються з: /р/ – загрозою та небезпекою; /л/ – лагідністю, ніжністю; /ш/ – тишею; /б/ – раптовістю, несподіваністю; /в/ – м'якістю, в'ялістю; /ж, з/ – енергійністю, жвавістю; /м, н/ – дієвістю; /к, п, с, т, ф/ – нейтральністю» [13, с. 338].

Як можна побачити, попри її, на перший погляд, примітивність, звуконаслідувальна лексика виявляє безліч цікавих ознак, проливає світло на питання ідіотетичності в людській мові.

1.2. Метамова дослідження

Щодо питання визначення звуконаслідування як такого, дослідники наводять різні формулювання, які, проте, концептуально мало відрізняються, однак різною є сама термінологія. Наприклад, О.Валігура дотримується такого визначення: «Звуконаслідування або ониматопею визначають як умовну словесну імітацію звучань навколишньої дійсності засобами певної мови» [3, с. 34]. І. Гаценко наводить таку дефініцію: «Термін звуконаслідування, або ониматопея (у процесуальному значенні), вживається на позначення безпосередньої звукової імітації звуків нелінгвістичної природи шляхом комбінації відповідних сегментів, які добираються з інвентарю фонем певної мови» [8, с. 132]. В.Охріменко окреслює поняття таким чином: «Ониматопея (від грец.: *ὀνοματοποιία*; *ὄνομα* – “ім’я”, *ποιία* – “творю”, тобто “словотворчість” через звуконаслідування), визначається як закономірний фонетично

мотивований зв'язок між звуковим складом слова (фонемами) й акустичною ознакою позначеного (денотатом)» [21, с. 604].

Відповідно до визначень науковців, по-перше, питомий український термін «звуконаслідування» та запозичений з грецької «ономатопея» вживаються синонімічно у всіх розглянутих дослідженнях, хоча більшість науковців частіше послуговуються терміном «ономатоп» на позначення одиниці звуконаслідувального класу лексики. По-друге, зазначається фонетично мотивований зв'язок між звуковим складом слова й денотатом. По-третє, ономатоп не тотожний вигуку.

1.3. Відмінність звуконаслідувальних слів та вигуків в українській мові

Вигуки є розрядом слів, які слугують для лаконічного вираження почуттів, волевиявлення та не співвідносяться з жодним реальним об'єктом позначення. Натомість звуконаслідування (ономатопи), хоч і мають обширний експресивний характер, співвідносні з певним об'єктом позначення, навіть якщо цей об'єкт є чимось нематеріальним [21, 8]. Наведемо характерні ознаки звуконаслідувань та вигуків у форматі порівняльної таблиці (табл. 1.1).

Таблиця 1.1. Порівняльні ознаки ономатопів та вигуків за І.Гаценко та В.Охріменко, узагальнені та доповнені власними міркуваннями.

Звуконаслідування / ономатопея / ономатоп	Вигук
Співвідноситься з явищами об'єктивної дійсності, які сприймаються одним з органів чуття людини (знак-ікона)	Безпосередньо виражає емоцію чи волевиявлення як знак-сигнал
Значення мотивується звуковим складом та може бути полісемічним	Значення є контекстуальним та характеризується омонімічністю

Продовження табл 1.1.

Може виконувати роль присудка, підмета, означення тощо	Дієслівно-вигуківі слова виступають лише у ролі присудка, а загалом вигуки не мають визначеної синтаксичної ролі у реченні
Не залежить від інтонації	Може залежати від інтонації
Є продуктивною лексичною групою, має індивідуальну варіативність	Багато в чому зберіг характер інстинктивного звука, коло вигуків досить обмежене
Значення асоціативно залежить від приголосних фонем, що можуть вказувати на якісні характеристики, такі як миттєвість, тривалість (якщо об'єктом позначення є дія або рух у просторі)	Важливішою є роль голосних фонем

Виходячи з наведених характеристик, можна також дійти висновку, що функціонування вигуків більше співвідноситься з фатичною та експресивною функціями мови, у той час як звуконаслідування є комунікативними одиницями, що несуть семантичну інформацію.

Переконливими є міркування про залежність значення вигуків від контексту у статті О.Мельник «Семантика вигука А в українській мові» [18], у якій на матеріалі 200 художніх творів українських авторів ХІХ–ХХІ століть було проаналізовано особливості вживання вигука А. Дослідницею було виокремлено 26 різних семантичних варіантів із 446 проаналізованих випадків вживання.

Інше дослідження, присвячене вигукам, провела Л. Шуст [24] на матеріалі творів Тараса Шевченка. Дослідниця зафіксувала понад 200 вигуків та поєднала

їх у 6 підгруп. Також виокремлено звуконаслідувальні слова, зазначаючи відмінність їхньої природи. У роботі Л. Шуст знаходимо підтвердження поданих вище міркувань: контекстуальну та інтонаційну зумовленість вигуків, яка проявляється повною мірою тільки в усному мовленні (напр. *Ох, Е-е, Ой* та ін.); використання повнозначних частин мови у ролі вигуків (напр. *Боже, Спасибі, Горе* та ін.) тощо.

Виходячи з усього вищезазначеного, вигуки будуть розглянуті як окремий клас лексики на практичному етапі роботи, однак більш докладна класифікація вигуків проводитися не буде, що дозволить більшою мірою сконцентруватися на звуконаслідуваннях.

Однак варто визнати, що існують й інші погляди на проблему. Можна згадати про дослідження Л.Мацько [17], яка ввела поняття «інтер'єктив», що об'єднує звуконаслідувальні слова та вигуки в один клас у межах звукозображувальної системи мови. Мовознавицею виокремлено такі семантичні класи вигуків: емоційні (*Ох, Ой, Ух, Ну, Отак*), імперативні (*Годі, Авжеж, Цить, Амінь*), вигуки мовного етикету (*Здоров, Слава богу, Добри вечір*) та інші.

Окрім того, автори словника-довідника «Українська мова» також пропонують відрізнити від звуконаслідувань та вигуків так звані «вербоїди», що функціонують як присудки в реченні. Як приклади таких одиниць наводяться: *стриб, хап, трах, шмиг, бубух*: «Коли це зміюка – бубух! Аж земля затряслась» (Укр. казки). Ми не будемо послуговуватися термінами «інтер'єктив» та «вербоїд» в нашому дослідженні, однак ці концепції є цікавими для розуміння різних поглядів на проблему.

1.4. Підходи до класифікації звуконаслідувальних слів в українській мові

Підходячи до питання класифікації ониматопів, потрібно взяти до уваги висновки, зроблені низкою різних дослідників.

Насамперед, звернемося до статті О.Валігури «Специфіка сучасних досліджень ониматопейчної лексики» [3], у якій подається поняття «звукозображувальної системи мови». Із визначення дослідниці випливає, що

під цим поняттям розуміється лексичне поле, одиниці якого охарактеризовані тим чи іншим чином як трансмісія фонетичною системою мови явищ об'єктивної дійсності через пряму асоціацію. За цією теорією, звукозображувальна система мови включає в себе звуконаслідувальну та звукосимволічну підсистеми. Перша покликана охопити звуконаслідування / ономатопи, тобто лексику, що має «акустичний денотат» (реальне нелінгвістичне вираження того чи іншого звукового явища). Натомість друга, звукосимволічна підсистема, охоплює лексику, що мотивована незвуковою ознакою, тобто передачею мовним звуком «відчуття, руху, враження від позначуваного предмета» [3, с. 32].

Клас звукосимволічної лексики є потенційно найбільшою відмінністю між звукозображувальною системою української та японської мови, адже у другій ономатопейчна лексика охоплює як слова з акустичним денотатом (під цим поняттям І. Гаценко [4] розуміє значення звуконаслідувань, що мотивоване явищами об'єктивної дійсності, що мають звуковий прояв), так і неакустичним (явищами об'єктивної дійсності, що мають інші прояви — візуальні, тактильні, емоційні), що буде продемонстровано в наступних розділах роботи.

Ми пропонуємо запелювати до визначеної О. Валігурою [3] звукозображувальної системи мови та окресили класифікацію, якою будемо послуговуватися в межах цієї роботи (рис. 1.1).



Рисунок 1.1. Система класифікації звукозображувальної системи мови на основі концепції О.Валігури [3].

Така типологія дозволяє нам розглядати звукозображувальної системи мови як таку, що включає в себе два класи одиниць зі схожою семантикою, однак низкою відмінних характеристик (розглянутих нами раніше). Ця концепція не виключає вигуки з поля розгляду звукозображувальної системи, визнано їхню подібність до звуконаслідувань, але дозволяє їх розрізняти. Це дає нам змогу сконцентрувати увагу виключно на класі ономатопейчної лексики в її двох проявах (звуконаслідувальна та звукосимволічна підсистеми).

Переходячи до класифікації власне ономатопейчної лексики, варто зазначити, що вона стосуватиметься переважно звуконаслідувальної підсистеми (а не звукосимволічної), у зв'язку з відсутністю укладених класифікації цієї групи.

У цьому аспекті у цілій низці праць І.Гаценко [8, 7, 5, 6] описано класифікацію звуконаслідувального класу лексики. Мовознавиця досліджувала звуконаслідування як у формально-структурному, так і семантичному аспекті.

Насамперед розглянемо типологію ономатопів за формальною ознакою. І.Гаценко [5] виділяє 3 великі групи ономатопів: 1) **однокомпонентні звуконаслідування**; 2) **двокомпонентні**; 3) **багатокомпонентні**.

Перша група включає в себе:

- односкладові звуконаслідування (*Туп, Ф'ю, Гуп*),
- двоскладові (*Кахи, Цвірінь, Пугу*).

Друга група охоплює ономатопи, що зазнали явищ

- **редуплікації** як у формі подвоєння односкладової основи (*Шарх-шарх, Хрум-хрум, Тьху-тьху*), так і двоскладової (*Шушу-шушу, Хаміль-хаміль*) та часткової редуплікації одного складу, що передає зміну якості звучання позначуваного явище (*Гугух, Бабах*).

- **парні ономатопи**, що мають або два різні односкладові / двоскладові елементи, кожен з яких є різним проявом самостійного явища (*Стук-грюк, Дзень-брязь*); або з двома компонентами — фонетичними варіантами один одного (*Тік-так, Біп-буп*) або з елементами, які є структурно зв'язаними й не

проявляють самостійної реалізації (*Куку-ріку, Туди-дενця*); або парні ономатопи з частковою редуплікацією (*Трах-тарарах, Шість-шелесть*).

- **ономатопи з єднальними сполучниками**, що мають значення тривалості або повторюваності (*Кахи та кахи, Хлясь да хлясь*).

Третя група включає багатокomпонентні ономатопи як з нерівнозначними, так і з однозначними основами (*Тра-ра-ра, Шу-шу-шу*). У рідкісних випадках фіксується неповна редуплікація основи з випаданням центрального приголосного (*Тра-та-та*) [5].

Можна зробити висновок, що кожен формальний клас ономатопів виражає різні акустичні особливості описуваного явища. Однокомпонентні звуконаслідування співвідносять з нетривалим, одиничним звучанням, не розтягненим у часі. Натомість явище повної редуплікації свідчить про багаторазовість у поєднанні з ритмічністю, незмінним повторюваним характером звучання. Багатокomпонентні та частково редупліковані нерівнозначні одиниці свідчать про переривчастість, вібрацію, подовження, зміну темпу, висоти та сили акустичної будови звуку.

У своїй дисертації «Типологічні особливості звуконаслідувальних слів» І.Гаценко також послуговується поняттями «анлаутних», «ауслаутних» та «інлаутних» фонем [7]. Науковиця не наводить експліцитного визначення цих термінів, однак індуктивно можна вивести, що анлаутною називається приголосна фонема у першій позиції у складі ономатопа або його редуплікованої форми. Відповідно ауслатна фонема займає останню позицію, а інлаутна — голосна фонема. Із цього стає зрозумілою інтерпретація ролі кожної фонери у складі ономатопа: «Голосні тонують звучання відповідно до своїх формантів, приголосні залежно від акустико-артикуляційних характеристик передають вид шуму» [7, с. 15].

Дослідницею також наведено кількісні дані стосовно частотності функціонування тих чи інших структурно-формальних типів визначених ономатопів в українській мові. Матеріалом цього дослідження слугували

дібрані звуконаслідування зі «Словника української мови в 11 т.» та «Словника української мови за ред. Б.Грінченка в 4 т.» (див. *табл. 1.2* та *табл. 1.3*)

Таблиця 1.2. Частка формальних груп ониматопів у реєстрі словників, визначена в дослідженні І.Гаценко [7].

Група	Частка із загальної вибірки ониматопів
Однокомпонентні звуконаслідування	60%
Двокомпонентні звуконаслідування	35%
Багатокомпонентні звуконаслідування	5%

Таблиця 1.3. Типи однокомпонентних ониматопів.

Тип структури	Частка із загальної вибірки однокомпонентних ониматопів
<i>ППП</i>	12%
<i>ПГ/Г:</i>	5%
<i>ПГС</i>	6%
<i>ПСПП</i>	13%
<i>ПСГС</i>	1%

*Умовні позначення: П — приголосний звук, Г — голосний звук,
: — подовження голосного.*

Визначення формально-структурних типів ониматопів слугує важливою теоретичною базою для проведення нашого дослідження завдяки потенціалу, який має ця класифікація для автоматизованої обробки ониматопейчної лексики.

Звертаючись до питання визначення тематичної класифікації ониматопів за семантичними ознаками, дослідники І. Гаценко, Л. Мацько [5, 17] виділяють загалом 4 великі класи (подаємо кількісні дані, визначені за принципом І.Гаценко [7]):

- Ониматопи, пов'язані з людиною та її діяльністю — 22% (наслідування звуків кашлю, ходьби, свисту, стуку зубів тощо);

- Ономатопи, що передають звучання явищ неживої природи та звуків рослин — 11% (звуки води, вітру, шелест дерев, грім тощо);
- Ономатопи, що наслідують звуки тварин та птахів — 24%;
- Ономатопи, пов'язані з предметами, які приводить у дію людина — 43% (звуки музичних інструментів, механізмів, предметів побуту тощо).

На нашу думку, ці класи є вичерпними, визначення семантичних груп в межах цих класів є доволі гнучким.

Можна також звернути увагу на факт деривативного потенціалу звуконаслідувань: вони можуть виступати базою для творення слів разюче інших частин мови. Наприклад, за допомогою іменникових суфіксів *-анн*, *-інн* («грюкання», «тупотіння», «гавкання») або дієслівних суфіксів, таких як *-а* («бряжчати», «цвірінькати», «бухикати»). Такі одиниці не входитимуть до поля розгляду нашого дослідження, адже вони є похідними та їхній деривативний зв'язок легко встановлюється з тим чи іншим ономатопом. Натомість варто визнати (про що свідчать і дослідження [4]), що звуконаслідувальні слова у своєму «чистому» можуть виступати в ролі різних членів речення, бути самостійною комунікативною одиницею (хоча, дійсно, зачасту тяжіють до предикативного елемента). Ці визначення стосуються синтаксичної та морфологічної ролі звуконаслідувань, що є цікавою темою для майбутніх досліджень — проте наразі нашою метою є розгляд насамперед семантичних категорій для побудови лексикографічної моделі.

Висновки до Розділу 1

Опис та класифікація звукозображувальної системи мови є проблемами, що відзначалися низкою українських науковців, таких як І. Гаценко [8, 7, 5, 6, 4], Л. Мацько [17], О. Валігура [3], які подають класифікації звуконаслідувань за різними принципами.

На прикладі досліджень нами було проаналізовано термінологію, якою послуговуються для опису ониматопейних (звукосимволічних) одиниць, визначено синонімічність термінів «звукосимволікування» / «ониматопея» / «ониматоп». Окреслено лексико-семантичні відмінності зазначених одиниць від вигуків. На основі досліджень визначено зміст понять «звукосимволівальна система мови», «звукосимволівальна підсистема», «звукосимволічна підсистема», «інтер'єктив» та інші.

Звукосимволівальну лексичну систему мови системи української мови, що об'єднує два класи одиниць — вигуки та звукосимволікування, що мають схожі та відмінні характеристики. Визначено більш докладну класифікацію звукосимволікувань, що охоплюють дві підсистеми — звукосимволівальну та звукосимволічну. Нами запропоновано узагальнену класифікацію одиниць в межах цих систем на основі зазначених робіт.

Розглянуто класифікацію українських ониматопів за двома принципами: формально-структурним та лексико-семантичним. Перший слугує теоретичною лінгвістичною основою для розуміння фонемного / графемного складу українських ониматопів. Другий слугуватиме як принцип класифікації одиниць у лексикографічному моделюванні, що буде описано в наступних розділах нашої роботи.

РОЗДІЛ 2. СТАНОВИЩЕ ОНОМАТОПЕЇЧНОЇ ЛЕКСИКИ В ДОСЛІДЖЕННЯХ З ЯПОНІСТИКИ

2.1. Дослідження японських онома топів українськими та японськими науковцями

Японську мову можна по праву охарактеризувати як таку, що є однією з найбагатших на звуконаслідувальну лексику серед мов світу — особливо якщо порівнювати з українською чи спорідненою з нею мовами. Про це свідчить низка досліджень японських лінгвістів та власний емпіричний досвід, пов'язаний з вивченням мови. Однак навіть на теренах країни вранішнього сонця ця тематика почала докладно розглядатися лише з другої половини ХХ століття, що пов'язано з особливостями японської мовознавчої традиції, яка до середини ХІХ століття фактично обмежувалася лише текстологічними дослідженнями — мовно-стилістичними описами класичних літературних текстів [15].

С. Уемура [40], наприклад, зауважував, що японську мову можна охарактеризувати одним словосполученням: «ごちゃまぜの言語» (гочямадзе но генго¹) — «змішана мова». Цим він хотів висловити той факт, наскільки багато іноземних впливів здійснено на мову, зокрема з боку китайської та т. зв. «австронезійських мов» [14]. Однак дуже характерним є те, що вже в структурі самого цього словосполучення наявний онома топейчний компонент: «ごちゃ» (гочя) — «мішанина / плутанина».

Велика увага до онома топейчної лексики є й наслідком її великого значення в японській мові), про що свідчать такі лексикографічні видання, як «Gendai Giongo Gitaigo Youhou Jiten», «Japanese-English Onomatopoeic Dictionary (Classification Type)», «Dictionary of Iconic Expressions in Japanese» [27, 25, 26]. Як зазначає О. Кобелянська [15], протягом останніх 40 років вийшли друком десятки як тлумачних, так і двомовних словників японської онома топейчної лексики, реєстри яких нараховують від кількох сотень до

¹ Тут та далі застосовано транскрипцію О. Коваленка для японської мови [16].

кількох тисяч ониматопів. Натомість у найґрунтовнішому дослідженні української звукозображувальної системи авторства І. Гаценко було нарахувано всього 363 ониматопи [7]. Наразі не укладено жодного цілісного і змістовного словника, в якому висвітлювалися особливості звуконаслідувальної системи саме української мови.

В українській японістиці питання ониматопів фактично досліджувалося лише в працях О. Кобелянської, які слугували для нас орієнтиром у цій частині теоретичного дослідження), торкалися цієї теми також проф. Ю. Мосенкіс [19] і Л. Богдан [1]. Також виданням, вартим нашої уваги й цінним для нашого дослідження, є «Японсько-український тематичний словник ониматопейчної лексики» Х. Егави та О. Кобелянської [9]. Про його структуру та зміст ітиметься в Розділі 3.

2.2. Звукосимволічність японської мовної традиції

У носіїв японської мови виявляється розширена символічно-асоціативна система з фонематичним складом своєї мови, про що зазначає С. Уемура [40], який посилається на численні експериментальні дослідження японських лінгвістів. Стверджується, що у більшості японців проявляється такі асоціації з відповідними приголосними фонемами: «/k/ — відчуття сухості; /s/ — відчуття приємності, вологості; /t/ — відчуття сили, чоловічого начала; /n/ — відчуття липкості, в'язкості; /h/ — відчуття легкості, відсутності опору (спротиву); вільне коливання, погойдування тощо; /m/ — відчуття круглості, жіночого начала; /y/ — відчуття м'якості, слабкості, податливості; /w/ — відчуття крихкості, непевності» [40, с. 14].

Якщо порівняти з відповідними ознаками, характерними для носіїв української мови (с. 11 нашої роботи), можна побачити суттєві відмінності у фоносемантичній інтерпретації фонем різних мов їхніми носіями не тільки серед самих асоціацій (та, звичайно, фонематичного складу самої мови), так і серед їхньої докладності (українці певну групу фонем характеризують просто як «нейтральні»). Однак є й низка подібних ознак, що зіставні з характеристиками, визначеними І. Гаценко, такими як: «звуки, що асоціюються

з ударами, зазвичай, передаються за допомогою ономатопоетичних коренів, що мають проривні приголосні; тривалі шуми передаються коренями ономатопоемом, що мають фрикативи; коливання, що поступово затухають, позначаються коренями з носовими приголосними» [7, с. 16].

Не можна також не зазначити про те, яку визначну роль відіграє сама фонетична система японської мови та її морфологія: в японській мові практично відсутні закриті склади, за винятком /n/, /m/ — єдиних фонем, які можуть мати місце наприкінці складу (однак навіть їх у японській мовній традиції вважають окремими морами — мінімально важливими одиницями, що відчуються психологічно й фонологічно, не дорівнюють ані звуку, ані складу [11]). Також абсолютно відсутні збіги будь-яких двох різних приголосних фонем, окрім подовжень. Відповідно до цього можна зробити висновок, що вся система японської мови та її звуконаслідувальна система характеризується більшою мірою вокалізмом, порівняно з українською, де жодне дослідження не наводило визначних асоціативних рис голосних фонем, а їхня роль зводилася до модуляції тональності описуваного звукового явища.

Стосовно символізму голосних фонем японський дослідник К. Таджіма [29] наводить такі асоціації: /a/ — характеризується «величністю, поширенням назовні»; /i/ — напруженістю; /u/ — приглушеністю, замкненістю; /e/ — «неабиякою забрудненістю»; /o/ — «заокругленістю, важкістю, спрямуванням всередину». Як приклад наводиться протиставлення двох ономатопів на позначення звуку грому: ガラガラ (*gara-gara*) та ゴロゴロ (*goro-goro*) — перший позначає звук «ніби грім падає з неба», а другий «звук грому в небі поміж дощових хмар».

Якщо врахувати, що в японській мові наявна обмежена кількість складів (всі можливі склади представлені в силабічних абетках хірагана/катакана з деякими доповненнями у вигляді складених палаталізованих складів — загалом їх близько 100), у японців також проявляється схильність до асоціювання цілих складів з тим чи іншим враженням чи явищем. Наприклад, згідно з М. Танно [31], し (*shi*) асоціюється з чимось тихим, げ (*ge*) — брудним, ふ (*fu*) — легким,

ㇰ (dze) — болючим, ㇰ (ny) — милим. Кореляцію також можна помітити між парами приголосним за дзвінкістю/глухістю та палаталізацією/відсутністю палаталізації: одзвінчення у складі ониматопів характеризує більшу гучність вираження явища, а палаталізація — м'якість, легкість сприйняття на слух [29]. Наприклад, とんとん (тон-тон) — Стук, どんどん (дон-дон) — Грюк, ぱりぱり (pari-pari) — Хрум, ぱりぱり (барі-барі) — Хрусть.

Специфічність японської асоціативної звукової системи була описана кількома експериментами. О. Кобелянська в одному зі своїх досліджень описує лінгвістичний експеримент, мета якого полягала у виявленні адекватного сприйняття семантики японської ониматопейчної лексики носіями української мови і навпаки. У результаті українські мовці зуміли дати 45% правильних відповідей, тоді як японці розпізнали семантику 65% звуконаслідувань, невідомої їм мови. Це свідчило про високу здатність носіїв мови з розвиненою звукозображувальною системою до побудови правильних асоціацій з невідомими їм звуковими поєднаннями, порівняно з мовцями, де ониматопейчна лексика є менш поширеною [16].

К. Таджіма також описує експеримент, проведений за участі мовців китайської та корейської мов, які мали відгадати значення 70 відібраних японських ониматопів та побудувати з ними речення (респонденти були знайомі з японською мовою на початковому або середньому рівні). Метою експерименту було встановити, наскільки важливу роль становить ониматопейчна лексика в лінгводидактичному аспекті, тобто чи достатнім є її висвітлення в навчальних матеріалах для іноземців та чи прозорі для них відповідні мотиваційні зв'язки. Попри прозорість усіх відібраних ониматопів для носіїв японської мови, іноземцям вдалося розпізнати лише семантику звуків тварин та деяких звуко символічних слів, які почасти включені до навчальних матеріалів. Загалом експеримент показав низький рівень асоціативної спроможності респондентів [29].

Як можна побачити, звуковий символізм відіграє надзвичайно важливу роль у світосприйнятті носіїв японської мови. Наступним етапом буде

визначення класів японської ономатопейчної лексики, аналогічно до принципів, викладених стосовно української мови в Розділі 1.

2.3. Структурно-формальна класифікація японських ономатопів

Форма звуконаслідувальних слів у японській мові відіграє важливу роль: саме за формою можна визначити, що те чи інше слово є ономатопом. Найвизначнішою рисою морфології японських ономатопів є редуплікація — абсолютна більшість таких одиниць мають саме подвійну структуру (усі приклади, наведені в попередньому пункті цього розділу, були саме такими). Редуплікація посилює образність і метафоричність звуконаслідувань, як зазначає О. Кобелянська [16]. І здебільшого саме форма є тим, що об'єднує ономатопейну лексику звуконаслідувальної та звукосимволічної підгруп — остання, як уже зазначалося, представлена в японській мові надзвичайно яскраво.

О. Кобелянська [16] наводить такий принцип узагальненої класифікації японських ономатопів за морфологічною структурою:

- нередупліканти — до цієї групи належать ономатопи, в яких відсутня редуплікація, натомість наявний або один корінь, або два нерівнозначні корені. Вона поділяється на:

1) Прості: прості непохідні (ずん (дзун) — глухий звук падіння); прості похідні, тобто такі, що утворені за допомогою додаткової афіксації と (то): がたん と (гата~~н~~то) — з гуркотом, різко), або り (рі): きらり (кіра~~рі~~) — яскравий блиск) або ん (н): ずしん (дзуші~~н~~) — важкий, глибокий звук або вібрація);

2) складні, що утворені двокореневою або трикореневою структурою з додаванням тим самих суфіксів: どたんばたん (дотан-бата~~н~~) — гуркіт, ふらりくらり (фура~~рі~~-кура~~рі~~) — хитко.

- Редупліканти — найбільша група ономатопів. Можуть існувати як без додавання афіксів: いそいそ (ісо-ісо) — жваво, так і з квазісуфіксом ん (н) в середині слова: ぶんぶん (бун-бун) — дзижчання, або квазісуфіксом り (рі): ふわりふわり (фува~~рі~~-фува~~рі~~) — плавно.

Проте найбільш вичерпну та всеохопну класифікацію ономатопів за

формою вираження наводить І. Таморі [30]. Повний список визначених груп ономатопів, прикладів та їхніх характеристик наведено в *табл. 2.1*.

Таблиця 2.1. Класифікація японських ономатопів за формою вираження (за І. Таморі [30]).

№	Фонемна структура	Приклади	Характеристика
1.	CVQ CVN CVV	かつ, ぱっ ばん, かん かー, ぐー	Односкладові ономатопи, що мають у якості ауслаутних компонентів гортанну змичку, <i>ん</i> (<i>n</i>) або подовження голосного.
2.	CVVQ CVVN	かーっ, ぱーっ かーん, がーん	Варіанти ономатопів з пункту 1 з подовженням голосного у випадку, коли ауслатною фонемою є гортанна змичка або <i>ん</i> (<i>n</i>).
3.	CVQ-CVQ CVN-CVN CVV-CVV	ぱっぱっ, ぶっぶっ ばんばん, かんかん かーかー, ぐーぐー	Редупліковані версії односкладових ономатопів з пункту 1.
4.	CVCVQ CVCVri CVCVN	ばたっ, ぼとっ ばたり, ぼとり ばたん, ぼとん	Двоскладові ономатопи, що мають у якості ауслаутних компонентів гортанну змичку, ^り (<i>ri</i>) або <i>ん</i> (<i>n</i>).
5.	CVQCVri CVNCVri	ばっさりどっきり ふんわりにんまり	Двоскладові ономатопи з інлаутною гортанною змичкою або <i>ん</i> (<i>n</i>) та ауслаутним компонентом ^り (<i>ri</i>).
6.	CVCV-CVCV	ばたばた, ぼとぼと どきどき, ころころ	Редупліканти двоскладових ономатопів. Найпоширеніша група.
7.	CVCVri-CVCVri CVCVN-CVCVN	ばたりばたり ばたんばたん	Варіанти редуплікованих ономатопів з пункту 6 з ауслаутним ^り (<i>ri</i>) або <i>ん</i> (<i>n</i>) для кожної з частин.

Умовні позначки: *C* — приголосний, *V* — голосний, *Q* — гортанна змичка (っ), *N* — суфікс *ん*, *ri* — суфікс ^り, дефіс — редуплікація.

Крім цього, зазначено також обмежену групу багатокomпонентних ономатопів, таких як こけこっこう (*кокекоккоо*), つんつるてん (*цунцирутен*) тощо.

Цікавим також є графічний аспект: ономатопи в складі японських речень почасти записуються як символами хірагани, так і катакани (двох японських силабічних абеток, що тотожні за змістом, але різні за формою та функціями). У разі якщо звуконаслідування записано катаканою, це слугує свого роду візуальним підсиленням, аналогічним до використання великих літер у європейських мовах, і свідчить про більшу інтенсивність вияву звукового явища, у той час як запис хіраганою є нейтральним або свідчить про м'якість та помірність.

2.4. Семантичні групи японських ономатопів

На відміну від української лінгвістики, в японській існує усталений чіткий поділ звуконаслідувань за семантичними ознаками, який включає 3 категорії: гісей-го (擬声語), гіон-го (擬音語) та гітай-го (擬態語) [15].

- Гісей-го, або голосозображувальні слова, — це група ономатопів, акустичним денотатом яких виступають голоси людей або тварин. Наприклад: きゃーきゃー (*кяа-кяа*) — високий жіночий крик; わんわん (*ван-ван*) — гав-гав.
- Гіон-го, або звукозображувальні слова, — ономатопи, акустичним денотатом яких виступає будь-який реальний звук природи або неживих предметів, дій, виконуваних людиною. Наприклад: さらさら (*сара-сара*) — звук водного потоку; ざーざー (*дзаа-дзаа*) — звук дощу під час сильної зливи.
- Гітай-го, або звукообразні (звукосимволічні) слова — особлива група слів, що описують явища з неакустичним денотатом, такі як візуальні враження, тактильні відчуття, емоції, способи дії тощо.

Наприклад: にやにや (*нія-нія*) — іронічно посміхатися; ふわふわ (*фува-фува*) — щось м'яке, пухнасте; ゆったり (*юттарі*) — стан розслабленості; ゆっくり (*юккурі*) — повільно тощо.

Згідно з І. Таморі [30], у вузькому значенні гітай-го не є ономатопами, однак серед японських лінгвістів більш поширене широке трактування, яке включає їх до складу ономатопейчної лексики. Логіка полягає насамперед у

формальній структурі цих слів: і гіон-го, і гісей-го, і гітай-го мають схожу (переважно редуplikовану) структуру, синтаксичну роль у реченні.

Відповідно до таксономії звукозображувальної системи української мови (звуконаслідувальна та звукосимволічна підсистеми), перші два типи можна співвіднести зі звуконаслідувальною підсистемою ониматопеїчної лексики (та частково з вигуками), а третій — з звукосимволічною підсистемою. Гітай-го також називають «міметичними словами» — ця група є визначною рисою японської мови, що не є характерною для багатьох інших мов (зокрема для української), однак, згідно з Х. Іносе, проявляється також в корейській та деяких африканських мовах [28].

Науковець також наводить розширену класифікацію японських ониматопів згідно до класів, описаних вище, використовуючи метод семантичного аналізу. Згідно з цією теорією, ониматопеїчна лексика поділяється на дві великі групи: слова, що мають сему «акустичне сприйняття» ([+AUDIBLE]), та слова, що її не мають ([-AUDIBLE]). Відповідно до цього гісей-го та гіон-го об'єднано в межах першої групи та протиставлено за семою «вираження голосу» ([+VOCALIZED]/[-VOCALIZED]), а щодо гітай-го подано більш докладну класифікацію відповідно до сем «ментальна діяльність» ([+MENTAL]/[-MENTAL]), «відчуття» ([+SENSATIONAL]/[-SENSATIONAL]) та «істота»/«неістота» ([+ANIMATE]/[-ANIMATE]) [30]. Розширену класифікацію за цим принципом представлено на *рис. 2.1*.

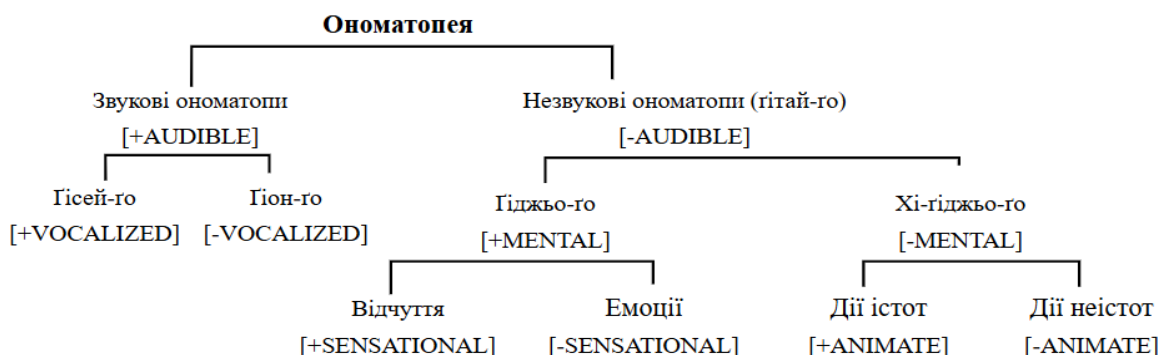


Рисунок 2.1. Класифікація японських ониматопів за семами, І. Таморі [30].

2.5. Проблеми японсько-українського перекладу

Звуконаслідування є дуже поширеним явищем переважно в японській

художній літературі, а тому слова цієї групи є неабияким викликом для перекладачів європейських мов, зокрема української. Як зазначає Д. Ржевська, «Японське світосприйняття відрізняється від європейського, і це має пряме відображення у лінгвістичних структурах. Японці сприймають світ “очима комахи”, а європейці – “з висоти пташиного польоту”, тому спосіб подачі японського тексту при перекладі вимагатиме ситуативних відповідників» [22, с. 216].

Класики японської літератури, поети, прозаїки, часто використовували гіон-го та гітай-го у своїх творах. Наприклад, зазначається, що Башьо Мацуо (XVII ст.) був одним із перших, хто почав вживати ономатопетичні слова в своїх віршах. Ця тенденція поступово ставала популярною [28]. Окрім того, ономатопетичні слова, які, як нами було вже з’ясовано, мають структуру з п’яти чи семи мор (якщо рахувати з граматичною часткою *と* (*to*), якою зазвичай ономатопеї вводяться в речення), дуже вдало вписуються в строфічні форми хайку й танка [16].

Згідно з О. Кобелянською [15], під час перекладу ономатопетичної лексики українською мовою у багатьох випадках необхідно використовувати прислівники, дієприслівники, прикметники чи дієслова, щоб найбільш точно передати їхню суть. За статистикою дослідниці визначає підбір функціонального еквівалента як такий, що використовується у 28% слововживань в перекладі, використання словникового відповідника — 12% та транскодування у 2% випадків.

Д. Ржевська [22], посилаючись на дослідження Х. Мінашіми, зазначає, що при перекладі японських ономатопів англійською мовою власне ономатопи фігурували в перекладі лише в 7,2% випадків. Подібну статистику на користь інших перекладацьких модуляцій наводить також Х. Іносе, посилаючись на дослідження перекладів шведською та іспанською мовами [28].

Наочним є приклад добору функціональних еквівалентів у перекладі японської народної казки (переклад І. Бондаренка «Качі-качі яма — Гора трісь-трісь» [2] (*табл 2.2*)).

Таблиця 2.2. Порівняння перекладу ономапонів в українському перекладі
 І. Бондаренка та оригіналі японської народної казки «Гора трісь-трісь» [2].

Переклад І. Бондаренка	Оригінал
Гора трісь-трісь	かちかち山
Що воно за звуки такі “клац-клац”?	今の「カチカチ」という音はなんだい？
Куме зайцю, а що воно за звуки такі “фу-фу”?	ウサギさん、この「ボウボウ」という音はなんだい？

Саме створення автоматизованого засобу для пошуку подібних словникових еквівалентів між двома мовами є метою нашого дослідження, у той час як особливості перекладацьких модуляцій та компенсацій не входить до розглядуваної проблематики та не є предметом створення лексикографічної моделі.

Насамперед також варто зазначити, що саме спосіб пошуку функціональних еквівалентів є основним в межах перекладу японської манги — виду японських коміксів. Манга є мультимодальним текстом, що послуговується ономатопами як фоновими дискурсивними елементами для кращої образності, «оживлення» тексту. Вони не займають багато місця, однак, своєю чергою, передають найменші відтінки того, що автор хотів би висловити, а також легко вписуються в малюнок, є своєрідними примітками, що дозволяють «почути» світ навколо літературних персонажів.

У такому випадку пошук лаконічних способів передачі сенсу є незамінним, і з власної практики відомо, що використання слів інших частин мови замість звуконаслідувань при такому перекладі є небажаним.

Висновки до Розділу 2

Проведено огляд досліджень ониматопеїчної лексики японської мови на основі праць українських та японських лінгвістів. Аналогічно до структури викладу матеріалу в розділі про дослідження звукозображувальної системи української мови, нами було визначено: 1) в яких аспектах розглядається ониматопеїчна лексика японської мови в лінгвістичних дослідженнях; 2) наскільки актуальною та поширеною є ця проблематика; 3) роль японських ониматопів у фоносемантиці японської мови та значення звукового символізму в лінгвокультурному / асоціативному аспекті; 4) докладну класифікацію японських ониматопів за формально-структурним показником; 5) таксономію японських ониматопів за семантичним показником, проаналізовано їхню семну структуру. Додатково було коротко означено проблематику перекладу ониматопів з японської на українську мову.

Найґрунтовніші дослідження в цій області проводили такі українські мовознавці, як О. Кобелянська [14, 15, 13, 9, 16], Д. Ржевська [22]. Найбільш яскраво ця проблематика висвітлена, звичайно, в роботах японських лінгвістів, таких як І. Таморі [30], К. Таджіма [29] та М. Танно [31].

Подібна структура підходу до питання ониматопеїчної лексики в двох розділах (Розділ 1 та Розділ 2) дозволяє нам в межах цієї роботи легше визначити спільні структурні й семантичні риси цих одиниць, спираючись на які буде проведено подальший аналіз та побудову лексикографічної моделі в наступних розділах нашого дослідження.

Враховуючи проблему перекладу, було також визначено роль та спрямування обраного методу пошуку функційних / словникових еквівалентів між японською та українською мовами, автоматизація якого і є нашою основною метою.

РОЗДІЛ 3. СТВОРЕННЯ БАЗИ ДАНИХ УКРАЇНСЬКИХ ОНОМАТОПІВ

3.1. Українська ономатопеїчна лексика: експертна оцінка електронних лексикографічних джерел

Перше завдання, що постало перед нами у межах практичної частини дослідження — це підготовка джерельної бази ономатопеїчної лексики. Як уже зазначалося, ця лексика не унормована у жодному окремому українському лексикографічному виданні, однак є можливість насамперед відібрати ономатопи із загальних тлумачних словників української мови.

І. Гаценко [7] проводила подібний добір звуконаслідувальної лексики на основі «Словника української мови в 11 т.» та «Словника української мови за ред. Б. Грінченка», у результаті чого дослідниця виокремила 363 ономатопи, однак не наводила повного реєстру цих одиниць.

Враховуючи той факт, що реєстр «Словника української мови в 11 т.» включено до більш повного «Словника української мови в 20 т.» (далі — СУМ-20), який до того ж подано у зручному електронному форматі на сайті [38], саме його було обрано як перше джерело для формування списку ономотопів. «Словник української мови за ред. Б. Грінченка» (далі — Словник Грінченка) [36] послугує другим джерелом для формування джерельної бази дослідження, однак у зв'язку з його перекладацькою природою (реєстр складають слова української мови, але визначення подаються російською мовою), такі словникові статті потребуватимуть перекладу.

Також для більш широкого охоплення буде розглянуто реєстр «Великого тлумачного словника сучасної української мови вид. “Перун”» (далі — Словник вид. «Перун») [32] на предмет наявності ономатопеїчної лексики, яка не трапилася у попередніх двох джерелах.

Усі три зазначені вище словники наявні в електронному форматі або у форматі сайту (СУМ-20), або у форматі оцифрованих текстових файлів (Словник Грінченка та Словник вид. «Перун»), що дозволяє автоматизувати процес добору словникових статей.

Крім цього, було вирішено розглянути й інші електронні джерела. Наприклад, «Великий електронний словник української мови» (далі — ВЕСУМ) [33] є зручним інструментом для виконання нашого завдання через його велику електронну базу слів, яким приписані певні теги (позначки), однак недоліком є відсутність словникових визначень, адже цей словник за своїм призначенням є граматичним. У такому разі визначення буде підібрано ситуативно з інших тлумачних словників, не зазначених вище, або написано вручну.

Ще одним корисним відкритим електронним джерелом для поповнення лексикографічної моделі є блог «livejournal» [59], на якому користувачем garhnaa було зібрано чимало специфічних ониматопейних слів, які ми вважаємо корисними для поповнення нашої бази. На сайті наведені слова з короткими визначеннями, категоріями та варіантами, що ми унормуємо та внесемо до бази даних.

Останнім, але не менш вагомим внеском до бази даних буде список ониматопейної лексики з доробку видавництва «Nasha Idea» [60], яке займається безпосередньо перекладом японської манги на українську мову. Цей список був наданий нам представниками видавництва у межах співпраці, усі слова фігурували у тих чи інших українськомовних виданнях та зібрані професіоналами у сфері японсько-українського перекладу.

Таким чином, джерельною базою для добору української ониматопейної лексики послугують:

- 1) СУМ-20;
- 2) Словник Грінченка;
- 3) Словник вид. «Перун»;
- 4) ВЕСУМ;
- 5) Блог «livejournal»;
- 6) Напрацювання видавництва «Nasha Idea».

Що стосується автоматизації процесу добору ониматопів з реєстру тлумачних словників, варто мати на увазі, що як звуконаслідування, так і вигуки

у них подаються системно з однією позначкою — «виг.» (або «меж.» у випадку Словника Грінченка). Як нами вже зазначалося в Розділі 1, вигуки та звуконаслідування ми розглядаємо як окремі класи слів у межах звукозображувальної системи мови, тому їхнє розрізнення буде здійснено вручну на основі їхніх визначень на подальших етапах роботи після їхнього добору. Вигуки не буде виключено з бази даних зовсім, однак їм буде приписано окрему категорію, яка дозволить провести розрізнення.

3.2. Екстракція української ономатопейчної лексики із електронних ресурсів та редагування бази даних

3.2.1. Словник української мови в 20 томах

Відповідно до списку джерел, представлених у попередньому пункті, першим етапом буде екстракція звуконаслідувальної (а заодно і вигукової, що є невіддільною в структурі словника) лексики із СУМ-20. Весь реєстр словника представлений на сайті sum20ua.com [38], і як раціональний спосіб добору словникових статей із сайту з такою структурою (рис. 3.1.) ми пропонуємо вебскрейпінг (за відсутності спеціального API функціоналу та наведеного офіційного повного реєстру словника в форматі бази даних).

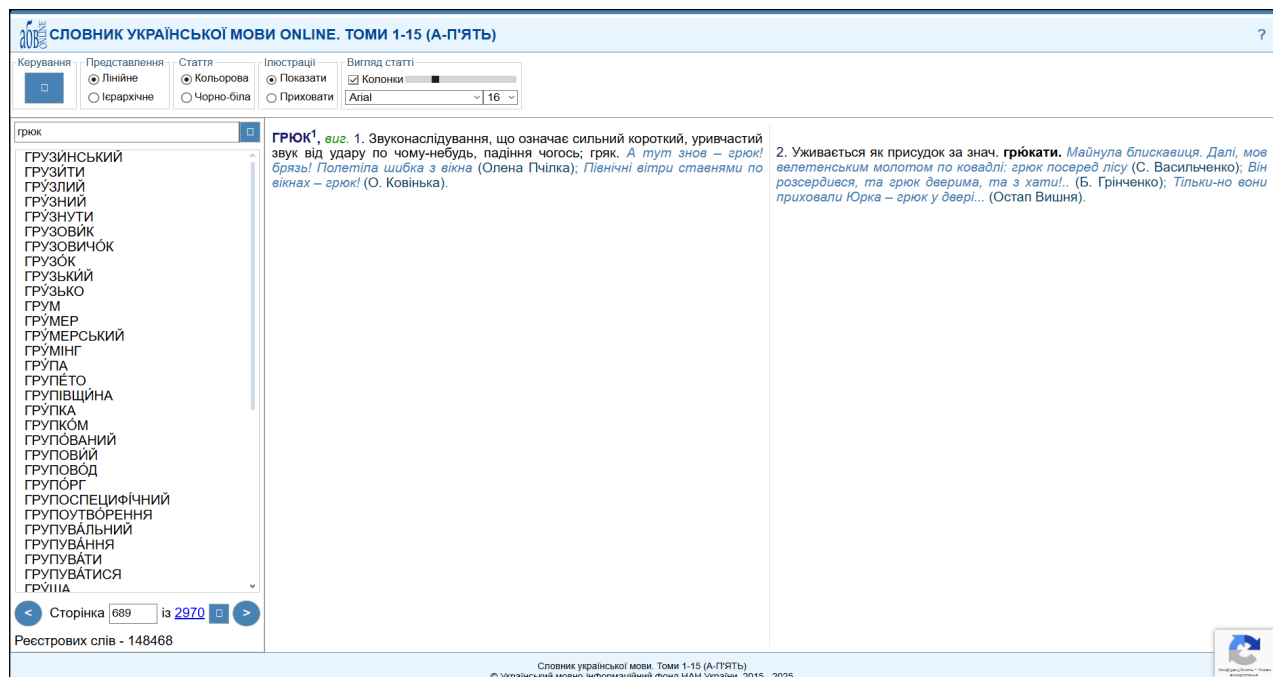


Рисунок 3.1. — приклад словникової статті в електронній версії СУМ-20.

Вебскрейпінг — це процес автоматичного збирання даних із вебсайтів за

допомогою спеціального програмного забезпечення. Вебскрейпінг дозволяє добирати та структурувати дані у формат, який можна зберігати й аналізувати в електронній таблиці або базі даних. Для проведення цієї процедури ми пропонуємо використати бібліотеку BeautifulSoup [42] на мові програмування Python [66]. Ця бібліотека є простою у використанні та, на нашу думку, здатна показати хороші результати, взявши за основу добре структуровані сторінки з розміткою HTML зі вказаного сайту.

Для екстракції одиниць із вебсайту було створено програмний код, який представлено у додатках через покликання на GitHub-репозитарій [Додаток 1]. Створення проєкту мовою Python [66] для добору словникових статей із СУМ-20 відбувалося у три етапи:

- 1) Вебскрейпінг посилань на всі словникові статті з сайту СУМ-20 (файл slovnyk20.py в репозитарії);

Використано бібліотеку requests [69] для завантаження HTML-вмісту сторінок словника. Бібліотеку BeautifulSoup с використано для парсингу та пошуку елементів у структурі сторінки. Код шукає список `<ul>` (у якому містяться посилання на всі статті в межах однієї сторінки), потім ітеративно дібрано всі `<a>`-посилання зі списку `<li>`, що містять `wordid` — тобто посилання на словникові статті. Цикл повторюється з переходом до наступних сторінок словника (їх 2970, на кожній переважно 50 словникових статей). Список усіх посилань зібрано у текстовому файлі `sum20ua_all_word_links.txt` (148468 словникових статей загалом).

```
<ul>
  <li>
    <a id="lid_20343"
      href="/?wordid=20343&page=688&searchWord=%D0%B3%D1%80%D1%8E%D0%BA#lid_20343">
        ГРУЗИНСЬКИЙ
    </a>
  </li>
```

Рисунок 3.2. — приклад подачі посилання на словникові статті в структурі HTML-сторінок онлайн-словника СУМ-20.

- 2) Відбір тих сторінок онлайн-словника, що містять скорочення «виг.» (файл vyhuk.py);

Використано асинхронні бібліотеки aiohttp [50] та asyncio [41] для одночасного отримання інформації з багатьох сторінок вебсайту та забезпечення ефективної паралельної обробки великої кількості сторінок. Ці етапи були необхідні для прискорення роботи програми.

Аналогічно до попереднього пункту використано бібліотеку BeautifulSoup [42] для аналізу HTML-вмісту сторінок і пошуку тих, що задовольняють наші умови (наявність позначки «виг.»). Функція fetch_and_parse() завантажує HTML-сторінки, парсить їх, шукає вміст з класом ENTRY, де наявний тег <i> з приміткою «виг.», та зберігає їх у файлі entries_with_vyg_async.txt.

```
<div>
  <div class="ENTRY"> <div class="WORD">ГРЮК<div class="UX">1</div></div><div class="LPART"><b>, </b> <i>виг.</i></div> <div
class="INTF"><div class="INTNUM">1</div>. <div class="FORMULA">Звуконаслідування, що означає сильний короткий, уривчастий
звук від удару по чому-небудь, падіння чогось; гряк</div>. <div class="ILL"> <div class="ILLTXT"> тут знов  грюк! брязь!
Полетіла шибка з вікна</div> <div class="ILLSRC">(Олена Пчілка)</div></div> <div class="ILL"> <div
class="ILLTXT">Північні вітри ставнями по вікнах  грюк!</div> <div class="ILLSRC">(Г. Ковінька)</div></div></div><div
class="INTN"><div class="INTNUM">2</div>. <div class="FORMULA">Уживається як присудок за знач. <b>грібкати.</b></div> <div
class="ILL"> <div class="ILLTXT">Майнула блискавиця. Далі, мов велетенським молотом по ковадлі: грюк посеред лісу</div>
<div class="ILLSRC">(Г. Васильченко)</div></div> <div class="ILL"> <div class="ILLTXT">Він розсердився, та грюк дверима,
та з хати!...</div> <div class="ILLSRC">(Б. Грінченко)</div></div> <div class="ILL"> <div class="ILLTXT">Тільки-но вони
приховали юрка  грюк  двері...</div> <div class="ILLSRC">(Остап Вишня)</div></div></div></div>
```

Рисунок 3.3. — Приклад структури словникової статті СУМ-20 в розмітці HTML.

3) Обробка отриманого HTML-файлу із словниковими статтями, які містять позначку «виг.», видобування лексичної інформацію зі структури HTML і зберігання її у базу даних SQLite [71] (файл sql_20t.py).

Використано sqlite3 [73], стандартну бібліотеку Python для роботи з вбудованими базами даних SQLite, та BeautifulSoup [42] для парсингу збережених попереднім кодом видобутих фрагментів HTML-розмітки. Проведено зчитування HTML, парсинг його структури, та збереження відповідних фрагментів до колонок бази даних: слово (word) — у блоці div class="WORD"; визначення (definition) — всі div class="FORMULA"; приклади (examples) — зібрано текст і джерело прикладу (ILLTXT, ILLSRC). Усі зібрані словникові статті збережені у базі даних vyg_dictionary.db з таблицею dictionary у трьох колонках: word, definition, examples.

	word	definition	examples
	Filter	Filter	Filter
22	Га-га	Звуконаслідування, що відтворює сміх.	1) У хату вскочив Матня. – Га-га, ...
23	Гав	Звуконаслідування, що відтворює ...	1) – Та не гавкай хоч ти [собако] ...
24	Гех	1) Звуконаслідування, яке відтворює ...	1) Підійме [дядько Микола] над ...
25	Ги-ги	Звуконаслідування, що виражає сміх.	1) Ця дурна дівка завжди продавала ...
26	Гойда-гойда	Звуконаслідування, що вживається як ...	1) Гойда-да! Гойда-да! Добра в коник...
27	Гоп	1) Звуконаслідування, яким ...	1) Сліпий по базару Оддирає ...
28	Гоц	Звуконаслідування, те саме, що "гоп".	1) Не кажи гоц, поки не перескочиш ...
29	Грим	1) Звуконаслідування, що означає ...	1) Стрільці роздмухують ґноти, ...
30	Грюк	1) Звуконаслідування, що означає ...	1) А тут знов – грюк! брязь! Полетіл...
31	Гу	1) Звуконаслідування, що передає ...	1) Гу-у! Оришко, гу! (М. ...

Рисунок 3.4. — Приклад записів у базі даних зібраних словникових статей з СУМ-20.

В результаті цього етапу роботи зі СУМ-20 було дібрано 172 одиниці. Більш докладний аналіз їхнього змісту (та даних з інших джерел) буде розглянуто в пункті 3.3. цього розділу.

3.2.2. Словник української мови за ред. Б. Грінченка

Оскільки реєстр СУМ-20 станом на травень 2025 р. опублікований лише у 15 томах (реєстр з літери А по П за алфавітом), лише його буде недостатньо для повного представлення жодного пласту лексики, однак його реєстр заклав фундамент для подальшого розширення нашої лексикографічної моделі. Наступним виданням, яке ми розглянемо, буде Словник Грінченка, що оцифрований та представлений у мережі [36]. За посиланням було завантажено словник у форматі *.pdf, і подальше його опрацювання включає такі етапи:

- 1) Конвертування файлу у форматі *.pdf у формат *.txt для програмної взаємодії з його змістом (використано ресурс convertio.co [44]) Словник збережено у файлі Hrinchenko_slovnyk.txt.
- 2) Відбір програмними засобами Python [66] всіх словникових статей, що містять скорочення «меж.» (файл grin.py):

Використано бібліотеку регулярних виразів re [68], за допомогою якої визначено межі словникових статей за формулою на *рис. 3.5*.

```
# Умова: рядок починається з великої літери і містить 'меж.'
if re.match(r'^[А-ЯҐЄІІА-Z]', line) and 'меж.' in line:
    current_article = [line.strip()]
    i += 1
    # Додаємо наступні рядки, поки вони не починаються з великої літери
    while i < len(lines):
        next_line = lines[i]
        if re.match(r'^[А-ЯҐЄІІА-Z]', next_line) and 'меж.' not in next_line:
            break
        current_article.append(next_line.strip())
        i += 1
    articles.append(" ".join(current_article))
```

Рисунок 3.5. — Логіка відбору словникових статей з текстового файлу оцифрованого Словника Грінченка.

Результати парсингу текстового файлу збережено в `Hrinchenko_mezh_filtered.txt`, а результати представлено на *рис. 3.6.*

Брику, меж. = Брик. Встрѣчено въ соедин. съ словомъ стріку: А кізочка стріку-брику. Рудч. Ск. I. 60.

Бринь! меж., выражающее звукъ струны. Бринь бандура та й замовкне. К. Досв.

Бруку! меж., выражающее воркованіе голубя. Глянь, як голубка в парі брукає: бруку- бруку, бруку-бруку, бруку-ку! Раулі. II. 115.

Брунь! меж. = Бринь. Зачепив бруньку за пакілець та брунь, брунь, смика її. Мнж. 109.

Бубух! меж. = Бебех! Преимущ. дѣтское. О. 1861. VIII. 8. Коли ж зміюка бубух! аж земля затряслась. ЗОЮР. II. 30.

Бугу! меж., выражающ. крикъ выпі. I бугай: бугу! гне з лози дугу. Лукаш. 85.

Буль-буль! меж., выражающее звукъ при вытеканіи жидкости изъ бутылки. Ном. № 13132.

Бульк! меж. Звукъ при паденіи въ воду. Мнж. 132. А жидівське військо в воду бульк! Грин. II. 223.

Бум-бум! меж., выражающее звукъ басовой трубы, бубна, барабана и т. п. О. 1862. VII. 74.

Рисунок 3.6. — структура відібраних словникових статей зі Словника Грінченка.

3) Парсинг текстового файлу, сформованого з вибірки словникових статей Словника Грінченка, які містять скорочення «меж.», зі збереженням даних у SQLite-базу даних (файл `grin_sql.py`).

Використано бібліотеку `sqlite3` [73] для створення бази даних SQLite та `re` [68] для парсингу слів, визначень та прикладів з текстового файлу `Hrinchenko_mezh_filtered.txt`. Статті в тексті розділені подвійним переносом рядка (`\n\n`), це дозволяє сформуванати список записів. Реєстрове слово співвідноситься з частиною тексту, що йде перед позначкою «меж.», визначення

— до першої крапки чи знаку оклику (якщо наведено кілька визначень з нумерацією, це також передбачено регулярним виразом). Решта тексту — відноситься до колонки прикладів. Логіка регулярних виразів для парсингу словникових статей представлена на *рис. 3.7*.

```
match = re.match(r"^(.*?)\s*меж\.\s*(.*)", entry.strip())
if not match:
    continue

word = match.group(1).strip()
rest = match.group(2).strip()

# Витягуємо визначення: до першої крапки або знака оклику
# Якщо після йде 2), 3) то включаємо ці значення теж
definition = ""
examples = ""
def_match = re.match(r"^(.+?[.!])(\s+(2\)|3\)|4\)|5\)|6\))?.*", rest)

if def_match:
    definition = def_match.group(1).strip()
    tail = def_match.group(2).strip()
    # якщо є ще підзначення, додаємо їх до визначення
    if re.match(r"^(2\)|3\)|4\)|5\)|6\))", tail):
        definition += " " + tail
        examples = rest[len(definition):].strip()
    else:
        examples = tail
else:
    definition = rest
    examples = ""
```

Рисунок 3.7. — Регулярні вирази для парсингу словникових статей у Словнику Грінченка.

На виході отримано базу даних `hrinchenko_mezh.db` з таблицею `dictionary`, що містить всі відібрані словникові статті з цього видання. Отримана таблиця потребує подальшого редагування, по-перше, через неповне дотримання системності у структурі словникових статей, і по-друге, у зв'язку з представленням визначень російською мовою (*рис. 3.8*). Для вирішення цього питання був необхідний ще один етап:

4) Автоматичний переклад словникових визначень з російської на українську (файл `grin_trans.py`).

Для автоматичного перекладу визначень з російської мови зі старою орфографією доцільним стало використання OpenAI API [64]. Кожне визначення було перекладено з таблиці `dictionary` та збережено в таблиці з

аналогічною структурою dictionary_translated в межах тієї самої бази даних hrinchenko_mezh.db. Повний код представлено у зазначеному файлі.

Ми визнаємо, що автоматичний переклад не може бути фінальним варіантом роботи, тому всі визначення були перевірені та відредаговані вручну в інтерфейсі бази даних через середовище DB Browser [72] (рис. 3.9). Автоматичний переклад слугує лише допоміжним інструментом для оптимізації та пришвидшення робочого процесу.

id	word		definition	examples
	Filter	Filter	Filter	Filter
1	1	A!	A, ах!	A, яке лихо! Впрочемь, чаще употребляется ой...
2	2	Абрюку,	, выражающее голубиное воркованіе.	Шейк.
3	3	Аву!	для выражения воя собаки, волка.	Шейк.
4	4	Авурр!	, выражающее ворчаніе собаки.	Прийшовши вона (лисичка) взяла ту торбу з ...
5	5	Ага-га,	1) Выражает удивленіе.	Ага-га, яка тут глибочинь! Шейк. 2) А, вотъ ...
6	6	Аги, агий, агі (на тебе)!	, выражающее порицаніе.	Желех. Агі на тя, пропав бис! Фр. Пр. 1.
7	7	Агов! Агось!	Означает призывъ и откликъ: Ау!	го, го! Эй! Послушай! Агоу, старий! — Агоу! ...
8	8	Агу!	1) = Агов. 2) Уже, вотъ, какъ вотъ. Агу, нашій ...	
9	9	Агу! Агучі!	Ласкательное обращеніе къ маленькимъ, ...	Агу, Івасю, агу, не плач!

Рисунок 3.8. — Приклад словниковий статей зі Словника Грінченка у базі даних до перекладу та редакції.

word	definition	examples
Filter	Filter	Filter
Брень	1) Звуконаслідування, що відтворює ...	
Брех	Звуконаслідування, що описує ...	Брех стара на покуті.
Бризъ	1) Звуконаслідування бризків або ...	Я ударив коня, а він бризъ об землю.
Брику-стрику	Звуконаслідування лягання або удару ...	А кізочка стріку-брику.
Бруку	Звуконаслідування, що виражає ...	Глянь, як голубка в парі брукає: ...
Брунь	1) Звуконаслідування, що відтворює ...	Зачепив бруньку за пакілець та брунь...
Бугу	Звуконаслідування крику птаха випи.	І бугай: бугу! гне з лози дугу.
Бум-бум	Звуконаслідування басової труби, ...	

Рисунок 3.9. — Приклад відредагованих словникових статей зі Словника Грінченка у базі даних.

В результаті цього етапу роботи зі Словника Грінченка було дібрано 414 нових одиниць з позначкою «меж.».

3.2.3. Великий тлумачний словник сучасної української мови вид. «Перун»

Наступним виданням, яке ми розглянемо, буде Словник вид. «Перун», що також оцифрований та представлений у мережі [32]. За посиланням було завантажено словник одразу в форматі *.txt, що спрощує задачу (файл Perun_slovnyk.txt). Для добору необхідних словникових статей з цього видання,

потрібно було виконати такий етап:

- 1) Відбір програмними засобами Python [66] всіх словникових статей, що містять скорочення «виг.» (файл perun.py).

Скрипт виконує фільтрацію словникових статей із текстового словника (у файлі Perun_slovnyk.txt) без імпорту додаткових бібліотек (словникові статті у файлі розділені двома переносами рядка \n\n, що дозволяє легко їх розпарсити), виділяє лише ті, що містять позначку «виг.». Результат записується у окремий файл Perun_interjections.txt.

Попри те, що відбір потрібних словникових статей з цього словника не викликав труднощів, складною є сама структура цих статей: наявні порушення кодування, що спотворює текст та безумовно потребує ручного редагування (рис. 3.10).

```
БРИЗЙСТИЙ, -а, -е, зах. Чорний з білими смугами.  
БРИК, виг. 1. Уживається як присудок за знач.  
брикАти 1 і брикйтися 1. 2. Уживається як при-  
судок за знач, брикнем.
```

Рисунок 3.10. — Приклад порушення кодування у файлі зі словником вид. «Перун».

До того ж у цьому виданні визначення, на нашу думку, є недостатньо розлогими (порівняно з попередніми джерелами), тому за можливості було проведено також їхнє доповнення (рис. 3.11).

```
Цю-цю - Вигук, що вживається з метою присоромлювання, висміювання когось.  
Шабаш - Вигук, що вживається як заклик закінчити щось або сигнал про закінчення чогось.  
Шаміль / Шаміль-шаміль - Звуконаслідування, що вживається для передачі обережного виходу, втечі кого-небудь звідкись.  
Шарах - Звуконаслідування, що використовується для позначення швидкої дії (за знач. "шарахнути", "шарахнутися" та ін.)
```

Рисунок 3.11. — Приклад виправлених та доповнених словникових статей зі Словника вид. «Перун».

Принципи доповнення визначень та їхнє внесення до бази даних розглянемо на прикладі наступного словника (ВЕСУМ), адже вони є

аналогічними.

Загалом зі Словника вид. «Перун» було дібрано 42 нові одиниці.

3.2.4. Великий електронний словник української мови

Згідно з інструкціями розробників ВЕСУМ, представленим на їхньому GitHub-репозитарії [33], було скачано файл `dict_corp_vis.txt` — словник у візуальному форматі (з відступами, згрупований за лемами) для перегляду, аналізу і опрацювання (рис. 3.12).

```
12  a-a intj
13  AAA noun:inanim:m:v_naz:nv:abbr    # кредитний рейтинг; тип батарейок
14    AAA noun:inanim:m:v_rod:nv:abbr
15    AAA noun:inanim:m:v_dav:nv:abbr
16    AAA noun:inanim:m:v_zna:nv:abbr
17    AAA noun:inanim:m:v_oru:nv:abbr
18    AAA noun:inanim:m:v_mis:nv:abbr
19    AAA noun:inanim:m:v_kly:nv:abbr
20  aaa intj:bad
21  a-a-a intj
22  a-a-a-a intj
23  аакуватий adj:m:v_naz
```

Рисунок 3.12. — Структура граматичного словника ВЕСУМ у файлі `dict_corp_vis.txt`

Завдяки простій структурі цього словника та наявності позначки «`intj`», можна легко вивести повний список наявних у ньому вигуків/звуконаслідувань:

Код у файлі `read_vesum.py` без використання додаткових бібліотек виводить у окремий текстовий файл `VESUM_intj.txt` повний список рядків, що мали позначку «`intj`» (з англ. *interjection* — вигук).

Оскільки ВЕСУМ не передбачає наявність визначень до слів, їх довелося збирати окремо (попередньо виключивши ті слова, що вже наявні в базі даних як результат попередніх етапів — за допомогою коду у файлі `VESUM_search.py`). Джерелом для цього слугували відкриті джерела, такі як вебсайт «Горох» [35], `slovnyk.ua` [39] (містять базу доданих користувачами слів). У разі якщо визначення не було знайдене, семантику слів було перевірено у контексті через корпус «ГРАК» [34] та написано самостійно (рис. 3.13).

Ава - Вигук на позначення несподіванки, здивування, шкодування і ін.
Боже - Вигук для вираження здивування, радості, страждання і т. ін.
Во - Вигук, що поєднується з жестом "палець вгору" та виражає одобрення, вподобання.
Кхе - Вигук, який використовується для привертання уваги як імітація кашлю.
Ла-ла / Ля-ля - Звуконаслідування, що вживається для відтворення мелодії.
М-м-м - Вигук, що виражає задоволення або іронічне ставлення.
Оп-ля - Вигук, що використовується перед стрибком або як заохочення до стрибка, фізичної дії.
О-ла-ла / О-ля-ля - Вигук, означає здивування, іронічне ставлення до чогось модного, незвичайного або як реакція на романтичну/інтимну ситуацію.
У-у-у - 1) Вигук, що використовується для вираження незадоволення, поєднується з жестом "палець донизу". 2) Звуконаслідування на позначення широкого кола явищ, подібних до гудіння, гулу, вітру або механічного шуму (напр. від холодильника).
Уф - Вигук, виражає втому, втому чи полегшення.
Тах - Звуконаслідування на позначення одноразового глухого пострілу.
Яволь - Вигук німецького походження, означає "так, звичайно", "звичайно", "слухаюсь!" або "так точно!".

Рисунок 3.13. — Приклад доповнених словникових статей з ВЕСУМ.

Відредагувавши та доповнивши словникові статті зі Словника вид. «Перун» та ВЕСУМ, збережені у файлах `Perun_interjections_edited.txt` та `not_found_VESUM_intj.txt` відповідно, їх можна імпортувати до бази даних: код у файлі `addition_VESUM_PERUN.py` дозволяє розпарсити текстові дані на реєстрову одиницю (`word`) та визначення (`definition`) та вставити ці дані до бази даних `combined_dictionary.db` таблиці `dictionary` зі вказанням джерела (`source`) — назви словника, з якого дібрано те чи інше слово.

Загалом з ВЕСУМ було дібрано 103 нові одиниці.

3.2.5. Інші електронні джерела

Останніми джерелами для поповнення нашої бази даних українських ономатопів, як зазначалося на початку розділу, стануть блог «livejournal» [59] та спеціально наданий нам робочий список звуконаслідувань з видавництва «Nasha Idea» [60]. У випадку з блогом (рис. 3.14) список був укладений у ручному форматі напряму з сайту, у зв'язку з невеликою кількістю таких одиниць, які, проте, є дуже цінними й цікавими.

ономатопея / звуконаслідування

Mar. 21st, 2011 at 1:43 PM



Ономатопея (грец. *ὀνοματοποιία*) — слово, що є результатом звуконаслідування. Найчастіше ономатопеїчною є лексика, яка прямо пов'язана з істотами або предметами — джерелами звуку: наприклад, дієслова типу "квакати", "нявкати", "кукурікати", "торохтіти" і похідні від них іменники... (wiki says)

звуконаслідування тварин:

баран бекає: бе-е; *ще*: ба-ба-ба; бе-ке-ке;

вівця мекає / бекає / бляє: ме-ке-ке;

вовк вис: у-у-у...; *ще*: в-а-у-у-у! *ще*: воу-воу;

ворона каркає: кар-кар; *ще*: (для передачі більш близького каркання) кра-кра;

гадюка сичить: сн-сн;

глухар токує: теке, теке, тек, тек, тек!

гогогун (вид чайки): го-го-го;

голуб вуркоче / буркоче / туркоче: буркуку;

горибей цвірінькає: цвінь-цвірінь / цвірінь-цвірінь / цінь-цвірінь / цінь-цінь;

грак/г(р)айворон гракає / г(р)аяє: гра-гра;

гусак гелкоче / гелгоче / гагає / гергоче / шайкає: га-га-га; *ще*: тра-тра-тра; га-га-га, гел-гел, ге-ге-ге;

джміль (політ): ж-ж-ж;

деркач: дер-дер;

жаба квакає / кумкає: ква-ква; кум-кум;

жайворонок співає: тью-і, тью-і, ті-і-і, трійю-тіх-тіх;

жук: ж-ж-ж;

журавель курличе: курли! курли! *ще*: кру-кру!

Рисунок 3.14. — Скріншот частини сторінки з блогу «livejournal» користувача *raphnana*.

Відібрані одиниці з визначеннями було збережено у файлі `from_blog.txt`. Загалом з блогу «livejournal» було дібрано 37 нових одиниць.

Наданий видавництвом «Nasha Idea» готовий список слів збережено у файлі `from_NI.txt`. — дібрано 111 нових одиниць.

Ці одиниці можна імпортувати до бази даних: код у файлі `addition_blog_NI.py` дозволяє розпарсити текстові дані на реєстрову одиницю (word) та визначення (definition) та вставити ці дані до бази даних `combined_dictionary.db` таблиці `dictionary` зі вказанням джерела (source) — назви словника, з якого дібрано те чи інше слово (аналогічно до коду в попередньому пункті).

3.3. Лінгвістичний аналіз бази даних української ономотопеїчної лексики

Підходячи до питання класифікації реєстрових слів на звуконаслідування та вигуки, ми керувалися такими принципами:

- 1) Якщо в оригінальному визначенні з того чи іншого словника зазначено напряму, що те чи інше слово є вигуком або звуконаслідуванням, його покласифіковано саме так.
- 2) Якщо визначення не було знайдено й було написано самостійно, до вигуків ми відносили одиниці, що мають характеристики, зазначені в класифікаціях І. Гаценко [8] та В. Охріменка [21] (докладно розглянуто в п. 1.3. Розділу 1).

Зазначимо, що це одиниці, які використовуються для вираження емоцій та водночас не співвідносяться зі звуковими явищами або іншими явищами об'єктивної фізичної дійсності (напр. *Ах*, *Гм*, *Ет*); одиниці, що виконують специфічну комунікативну або фатичну функцію, таку як привітання або звертання уваги (напр. *Алло*, *Турай*, *Здрастуй*), у тому числі вигуки для призову / відгону тварин (напр. *Киш*, *Цюцю*, *Тпру*).

Оскільки вигуки не є основним об'єктом інтересу в межах цього дослідження, більш докладна класифікація вигуків на підгрупи не проводилась. Стосовно звуконаслідувань, було розроблено додаткову класифікацію звуконаслідувань на підгрупи відповідно до джерела походження того чи іншого звукового (візуального або тактильного) явища (рис. 3.15). Ця класифікація подібна до класифікації, розробленої І. Гаценко [7] та розглянута нами докладно в п. 1.4. Розділу 1.

Звуконаслідування представлені у п'ятих великих групах: 1) Фізіологія людини, звуки, що пов'язані з тією чи іншою частиною тіла; 2) Звуки, що пов'язані з голосами або життєдіяльністю тварин; 3) Звуки, що пов'язані наляму або опосередковано з тим чи іншим проявом діяльності людини, переміщенням у просторі або взаємодією з певними об'єктами; 4) Звуки неживої природи; 5) Інші специфічні звукові прояви.

Загальна кількість підгруп в межах цих категорій — 36.

Звуконаслідування				
Людина	Тварини	Діяльність людини	Природа	Інше
Кашель	Птахи	Рух	Вітер	Магія
Сміх	Звірі	Вибух	Рідина	Емоції (звуки, пов'язані з проявом емоцій, а не відповідні вигукі)
Їда (пережовування їжі, ковтання тощо)	Комахи	Удар	Світло	Тиша
Свист	Інші тварини	Постріл	Погода (дощ, грім тощо)	Шум (всі інші види звуків, що не підпадають під характеристики інших категорій)
Рот (скрежет зубів, поцілунки тощо)		Стан		
Дефекація		Музика		
Сон		Техніка		
Дихання		Дзвін		
Розмова				
Очі				
Плач				
Дія рукою				
Крик				
Ніс				
Стрибок				
Ходьба				

Рисунок 3.15. — Класифікація звуконаслідувань на підгрупи за семантичним принципом (авторська розробка).

Мотивація до призначення того чи іншого класу словам впливає переважно з самих дефініцій, в яких зазначені поняття з тих самих семантичних полів.

Окремого пояснення потребує група «емоції», що нами була виокремлена в межах звуконаслідувальної лексики, попри те, що прояв емоцій є також характеристикою вигуків. Ця незначна група включає одиниці, що є своєрідним опосередкованим звуковим символізмом емоцій, на противагу їхньому безпосередньому вираженню. Наприклад: *Вереду* — вживається на позначення вередування дитини (у скоромовках), *Тьох* — виражає емоційні переживання (як-от швидке биття серця), *Тін* — виражає нерве сіпання, роздратованість.

Варто також звернути на увагу на те, що було виявлено 2 одиниці, що були класифіковані як окремий специфічний тип — фразеологічно зв'язаний вигук. Слова *Куць* та *Фай* вживаються лише в складі відповідних

фразеологічних виразів: «*Куць виграв, куць програв*» — про мінливість подій; «*Фай би му!*» — прокляття або побажання лиха. Такі словникові статті були виявлені в реєстрі Словника Грінченка.

В результаті етапів добору словникових статей можемо навести таку статистику (табл. 3.1).

Таблиця 3.1. — Статистика результатів наповнення бази даних.

Джерело	Сумарна кількість	Кількість звуконаслідувань	Кількість вигуків
СУМ-20	172	94	78
Словник Грінченка	414	265	147 (+2 фразеологічно зв'язані вигуки)
Словник вид. «Перун»	42	13	29
ВЕСУМ	103	47	56
Блог «livejournal»	37	33	4
Доробок вид. «Nasha Idea»	111	109	2
Всього:	879	563	316

Якщо порівняти ці результати з дослідженням І. Гаценко [7], нам вдалося зібрати на 55% більший реєстр ономатопів, ніж той, що розглядала дослідниця (563 проти 363 одиниць).

Деякі одиниці, що мали аналогічні визначення та подібну форму, були об'єднані в один рядок як варіанти (через символ /). Якщо порахувати всі варіанти, то кількість відібраних звуконаслідувальних слів складатиме 712 одиниць (приріст 96%, порівняно з дослідженням І. Гаценко).

Повна таблиця прокласифікованих словникових статей доступна за посиланням в Додатку 3.

Висновки до Розділу 3

У результаті проведеної роботи було проаналізовано вміст шести джерел для добору матеріалів поповнення бази даних української ономотопеїчної лексики, серед яких: «Словник української мови в 20 т.» [38], «Словник української мови за ред. Б. Грінченка» [36], «Великий тлумачний словник сучасної української мови вид. “Перун”» [32], «Великий електронний словник української мови» [33], блог «livejournal» [59], список ономотопеїчної лексики з доробку видавництва «Nasha Idea» [60].

Було використано засоби мови програмування Python [66] для автоматизованої обробки вмісту вказаних джерел та їхнього внесення до реляційної бази даних. Використано такі програмні бібліотеки: sqlite3 [73], re [68], BeautifulSoup [42], aiohttp [50], asyncio [41].

Для автоматичного перекладу словникових статей «Словника української мови за ред. Б. Грінченка» з російської на українську застосовано програмні засоби OpenAI API [64], після чого в інтерфейсі програми DB Browser [72] проведено редагування вручну.

Усі здобутки розділу об'єднані та збережені у базі даних файлу combined_dictionary.db [Додаток 2] таблиці merged_ukrainian_onomatopoeia, де словникові статті представлені за такими полями: word (реєстрове слово), definition (визначення), examples (приклади вживання), source (одне з вказаних вище шість джерел), category (приписана семантична категорія).

Проведено лінгвістичний аналіз бази даних, реєстр прокласифіковано на вигуки та звуконаслідування, у межах звуконаслідувань виокремлено п'ять груп, що поділяються на тридцять шість підгруп.

Загальна кількість відібраних слів з шести різних джерел становить 879 одиниць без урахування варіантів (слів, що мають однакове визначення і функціонування та відрізняються за формою не більше ніж на дві графеми або є відповідними редуплікантами), що об'єднані вручну в межах однієї словникової статті, якщо такі наявні.

РОЗДІЛ 4. СТВОРЕННЯ БАЗИ ДАНИХ ЯПОНСЬКИХ ОНОМАТОПІВ

4.1. Японська ономатопейна лексика: експертна оцінка електронних лексикографічних джерел

Що стосується японської мови, як вже зазначалося, ономатопейна лексика в ній представлена більш широко, ніж в українській, про що свідчили дослідження, розглянуті в Розділі 2 [12, 13, 14, 15, 16, 22, 29]. Нами також було з'ясовано, що в межах японського мовознавства було укладено чимало спеціалізованих лексикографічних видань для її ілюстрації [26, 27]. В межах нашого дослідження ми маємо на меті укласти словник японських ономатопів для їхнього перекладу українською (поєднання української та японської баз даних буде розглянуто в Розділі 5 нашої роботи), з чого випливає перша бажана умова відбору джерел: наявність українських визначень до японських ономатопів. Існує лише одне видання, що пропонує саме такий зміст — «Японсько-український тематичний словник ономатопейної лексики» Х. Егави та О. Кобелянської [9]. Автори зазначають, що словник містить близько 500 ключових і понад 2 тис. похідних ономатопейних слів сучасної японської мови, згрупованих за тематичним принципом (семантичними категоріями). Цей словник слугуватиме першим джерелом для поповнення нашої лексикографічної бази даних.

Беручи до уваги широке висвітлення ономатопейного пласту японської мови як в академічних виданнях, так і в інтернеті та автоматизований формат нашого практичного завдання, другою умовою відбору стане електронне представлення таких словникових даних. У мережі наявна чимала кількість електронних оцифрованих ресурсів, що можуть слугувати матеріалом для виконання нашого завдання, насамперед, ми пропонуємо розглянути:

- 1) Базу даних ресурсу nihongoresources.com (далі — Словник «nihongoresources») [61], що представлена на GitHub-репозитарії розробників [62]. Цей словник містить понад 3 тис. статей, представлених у зручному для автоматизованого опрацювання форматі.

2) Вільний користувацький електронний словник «Japanese Manga SFX» (далі — Словник «SFX») [57], що містить близько 2 тис. статей, він наповнений звуконаслідуваннями, що реально трапляються в текстах японської манги.

Обсягу реєстрових одиниць, наявних в цих трьох джерелах, на нашу думку, буде більш ніж достатньо для адекватного охоплення звуконаслідувальної лексики, що підлягає еквівалентному перекладу українською (враховуючи, звісно, той факт, що японські словники будуть містити широкий пласт лексики гітай-го — звукосимволічних / міметичних слів, що не піддаються еквівалентному перекладу звуконаслідуваннями української мови. Їхню класифікацію буде проведено в наступних етапах роботи).

Недоліком для нас є те, що визначення в словниках «nihongoresources» та «SFX» представлені англійською мовою. У зв'язку з обсягом роботи (понад 5 тис. словникових статей з цих двох джерел), аналогічно до опрацювання Словника Грінченка в Розділі 3, ми скористаємося засобами автоматичного перекладу для оптимізації роботи, після чого всі визначення будуть переглянуті й відредаговані вручну для отримання адекватних результатів.

4.2. Екстракція японської ономатопейчної лексики із електронних ресурсів та редагування бази даних

4.2.1. Японсько-український тематичний словник ономатопейчної лексики

Це лексикографічне видання, як зазначається авторами, «розраховано на перекладачів, студентів та викладачів вищих навчальних закладів України» [37, с. 17]. Воно є єдиним спеціальним словником ономатопейчної лексики, його специфіка тісно пов'язана з тематикою нашого дослідження. Однак попри безсумнівну цінність цієї роботи, приклади еквівалентного перекладу японських ономатопів українськими хоч і представлені, однак, на нашу думку, потребують суттєвого доповнення.

Реєстр словника побудований за тематичним принципом та представлений у межах п'яти категорій: фізичні явища, природні явища, живі істоти та рослини, людина, суспільне життя — кожна категорія також містить підкатегорії, сумарна кількість яких становить 33. Словникові статті

структуровано за таким принципом (рис. 4.1): реєстрове слово хіраганою (ліва колонка); транскрипція ромаджі; частина мови, у ролі якої функціонує слово; визначення; частки та дієслова, з якими слово часто використовується; приклади японською мовою; переклади прикладів українською мовою; близькі за значенням слова (права колонка).

ひらり	[hirari] прислівник Перевертатися з легкістю, розгортатися (про одноразове перевертання чи розгортання легкого і тонкого предмета /листя, пелюстків, тонкої тканини, паперу тощо/). 〜と、ひるがえる 国旗^{こっけい}をひらりとひるがえして、選手^{せんしゅ}がスタジアム^{にゅうじょう}に入場してきます。 Одним рухом легко розгорнувши національний прапор, на стадіон увійшли гравці. Близьке за значенням: ひらっ
-----	--

Рисунок 4.1. — Приклад словникової статті з «Японсько-українського тематичного словника ономатопейчної лексики» [17].

Словник представлений у мережі в оцифрованому форматі за посиланням [37], що дозволяє включити його до розгляду в межах цієї частини роботи та провести його автоматизовану обробку і включення його реєстру до бази даних.

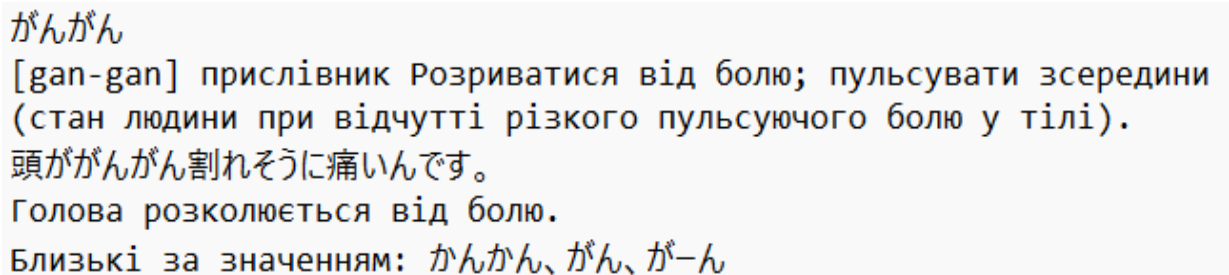
Для цього було виконано такі етапи:

- 1) Завантажений файл у форматі *.pdf ми конвертували у формат *.txt для програмної взаємодії з його змістом (використано ресурс convertio.co [44])
Словник збережено у файлі Yaponsko-ukrainskyi_tematychnyi_slovyk_onomatopeichnoi_leksyky.txt.
- 2) Стикнувшись із проблемами, пов'язаними з коректністю конвертації файлу в текстовий формат, нами був розроблений скрипт мовою програмування Python (код у файлі japanese_dict_clean.py) для виправлення неточностей та приведення тексту до формату, який можна буде автоматизовано внести до

бази даних SQLite:

Використано бібліотеку регулярних виразів `re` [68], за допомогою якої виконано такі етапи обробки: видалено подвійні переноси рядків; номери сторінок словника; символи фурігана (символи японської силабічної абетки хірагана, що використовуються як надрядкові елементи для транскрипції ієрогліфів канджі — у цьому випадку, в лінгводидактичних цілях. Адекватне конвертування таких транскрипцій у текстовий формат неможливе, проте вони є опціональними та не впливають на зміст); зведено в один рядок розбиті визначення українською мовою та приклади японською мовою в межах кожної словникової статті. Результат автоматизованої обробки збережено у файлі `cleaned_step2_final_all_filtered.txt`.

- 3) Для усунення несистемних помилок у конвертування файлу, було виконано ручне редагування тексту в файлі, отриманому на попередньому етапі. Усі словникові статті були зведені до формату, як на *рис. 4.2*. (у кожному окремому рядку: реєстрове слово, визначення, приклади японською, переклади прикладів українською, близькі за значенням слова) та збережено у файлі `reordered_cleaned_text.txt`.



がんがん
[gan-gan] прислівник Розриватися від болю; пульсувати зсередини
(стан людини при відчутті різкого пульсуючого болю у тілі).
頭ががんがん割れそうに痛いんです。
Голова розколюється від болю.
Близькі за значенням: かんかん、がん、がーん

Рисунок 4.2. — Приклад проміжного редагування словникових статей у файлі до внесення в базу даних.

- 4) Відредагувавши словникові статті зі Словника, їх можна імпортувати до бази даних: код у файлі `add_jp_egawa.py` [Додаток 1] дозволяє розпарсити текстові дані на реєстрову одиницю (`hiragana`), визначення (`definition`), японські приклади (`examples`), переклади прикладів (`examples_translation`), близькі за значенням слова (`similar`) та вставити ці дані за допомогою засобів `sqlite3` [73], стандартної бібліотеки Python для роботи з вбудованими базами

даних SQLite, до бази даних combined_dictionary.db таблиці jp_onomatopoeia з відповідною структурою. Інтерфейс бази даних представлено на *рис. 4.3*. (переглянуто через середовище програми DB Browser [72]).

hiragana	definition	examples	examples_translation	similar
あいあい	Спокойно, тихо, ...	結婚したら、和気あいあ...	Після одруження ...	NULL
あくせく	Одержимо, помішано...	会社で、毎日あくせく働...	Працюю, мов ...	NULL
あたふた	Метушливо, ...	1) 突然の質問にあたふ...	1) Розхвилювався ві...	NULL
あつ	Дуже короткий ...	あつという間に、一年もた...	Не встиг і ахнути ...	NULL
あつ	Вигук "Ой!"	1) アッ、どうしよう。財布...	1) Ой! Що робити? ...	Влизькі за значенням: あつ、あー、あれっ
あつさり	Просто, легко, без...	1) 問題をあつさり片付け...	1) Легко вирішити ...	Влизькі за значенням: さつ、さっさ、さらっ
あつさり	Нехитрий, ...	あの人は口は悪いが、あ...	Та людина хоч і ...	Влизькі за значенням: からっ、さらっ
あつさり	Легкий, ...	日本料理はウクライナ料...	Страви японської ...	Влизьке за значенням: さらっ
あっぱあっぱ	Буль-буль! ...	うまく泳げないので、プー...	BLOV	NULL
あはは	Хи-хи-хи! Ха-ха-ха...	あはは、面白い笑い話で...	Хи-хи-хи! (Ха-ха-...	Влизькі за значенням: ははは、へへへ、ほほほ
ありあり	Виразно, чітко, ...	10年経ったのに、あのこ...	Минуло вже десять ...	Влизьке за значенням: まざまざ

Рисунок 4.3. — Структура бази даних із внесеними словниковими статтями «Японсько-українського тематичного словника ономатопейної лексики»

В результаті зі словника було дібрано 796 словникових статей.

4.2.2. База даних ресурсу «nihongoresources»

Розробниками цього ресурсу на їхньому GitHub-репозитарії [62] представлено базу даних японських ономатопей у текстовому форматі. Було скачано файл giongo.txt — словник, що містить 4 поля: хірагана, катакана (альтернативний запис реєстрового слова засобами другої японської силабічної абетки), коротке визначення англійською та призначена категорія (сміх, мовлення, дія, рух тощо). (*рис. 4.4*).

hiragana	katakana	translation	applies to	more specifically
あ	ア	""uh"", ""ah"", ""oh""	manga sound effects	
あー	アー	""yes"", ""okay""	manga sound effects	
あーあー	アーアー	yawning onomatopoeia	human noises	
あーっ	アーツ	strong version of あつ	manga sound effects	
あーん	アーン	crying with the mouth open	emotions	crying
あーん	アーン	opening (the mouth) wide	manga sound effects	
あーん	アーン	crying onomatopoeia	human noises	
あーんあーん	アーンアーン	crying onomatopoeia	human noises	
あぎあぎ	アギアギ	"bite, gnaw"	manga sound effects	
あくあく	アグアグ	bite	manga sound effects	
あくせく	アクセク	"toil, drudge, restlessly"	work attitude	
あたふた	アタフタ	hurriedly	action	
あたふた	アタフタ	"flustered, panicky"	mental state	

Рисунок 4.4. — Структура файлу giongo.txt від nihongoresources

Завдяки чіткій структурі файлу (кожне поле відділене одне від одного табуляцією \t, а кожна словникова стаття представлена в окремому рядку), його вміст можна одразу імпортувати до бази даних: код у файлі `add_japanese.py` [Додаток 1] дозволяє розпарсити текстові дані на реєстрову одиницю (hiragana), запис катаканою (katakana), визначення (definition), категорія (category) та вставити ці дані за допомогою засобів `sqlite3` [73], стандартної бібліотеки Python для роботи з вбудованими базами даних SQLite, до бази даних `combined_dictionary.db` таблиці `japanese_onomatopoeia` з відповідною структурою. Приклад структури бази даних представлено на *рис. 4.5*.

hiragana ▲	katakana	definition	category
Filter	Filter	Filter	Filter
からんからん	カランカラン	"hollow ""clang"""	onomatopoeia bells
がし	ガシ	grab	manga sound effects
がし	ガシ	"to hold tightly, to hug tightly"	performing an action
がっしり	ガッシリ	solid	manga sound effects
がっしり	ガッシリ	solidly built	resultant state
がっしり	ガッシリ	"well-built, solid, muscular"	the body
がっははは ...	ガッハハハ ...	evil laughter	manga sound effects
がっぶり	ガツブリ	grip tightly	action
がっぶり	ガツブリ	to bite down	performing an action
がつつがつつ	ガツツガツツ	with a sharp heavy object	performing an action stabbing
がぶっ	ガブツ	swallow in one gulp	performing an action drinking
がぶっ	ガブツ	bite something in one chomp (focus o...	performing an action
がぶり	ガブリ	swallow in one gulp (slower than が...	performing an action drinking

Рисунок 4.5. — Приклад структури бази даних з внесеними словниковими статтями nihongoresources до перекладу.

Переклад словникових визначень ми розглянемо разом із даними з наступного ресурсу, адже процес є аналогічним.

4.2.3. Користувацький електронний словник «Japanese Manga SFX»

Словник SFX [57] є вільним, тобто наповнений користувачами, а не науковцями, проте рівень його структурованості та обсяг реєстру є дуже високими, що зумовлює наш вибір саме його як одного з джерел в межах нашого дослідження.

На зазначеному сайті [57] словник представлений у форматі електронної

таблиці GoogleSheets, слова зібрані за алфавітом (порядком японської силабічної абетки хірагана), представлені записи хіраганою, катаканою, транскрипцією ромаджі, приклади або визначенні англійською мовою та короткі пояснення до більшості слів (рис. 4.6).

ア			
Japanese	Romaji*	English	Explanation
ア, あ	a	*whoa*	
アッ, あっ	a	ah	exclamation of surprise, amazement, relief
ア!, あ!	a	*hey!* *ah!*	SFX for a crying or calling out.
ア。。。, あ。。。	a	*ah...* *ah* *oh* *ahh*	Kind of like an open-mouthed "oh, I see..."

Рисунок 4.6. — Приклад структури електронної таблиці Словника SFX [57].

Для автоматизованої обробки даних цього джерела було виконано такі етапи:

- 1) Напрямку з сайту завантажено версію таблиці в розширенні *.csv (файл JapaneseMangaSFX-Sheet1.csv).
- 2) Написано код на Python (файл SFX.py у нашому репозитарії) для автоматизованої обробки та переформатування даних у формат, що буде підходити для імпорту до бази даних.

Використано такі бібліотеки: csv [46] (базова бібліотека Python для зчитування вмісту CSV-файлу), itertools [56] (для пропуску перших двох рядків-заголовків у CSV-файлі під час обробки.), dataclasses [47] (для оголошення простої структури даних RawEntry без потреби вручну писати

конструктор `__init__`), `typing` [75] (для анотації типу `details` як необов'язкового у `RawEntry`), `re` [68] (для обробки та нормалізації текстових рядків за допомогою регулярних виразів), `collections` [43] (для групування записів за катаканною, автоматично створюючи список для кожного ключа), `json` [58] (для запису фінального результату у форматі JSON). Основна функція коду `compile_japanese_manga_sfx_sheet(entries)` покликана проводити зчитування файлу з двома заголовковими рядками, визначати колонки, які містять японські слова, ітерувати по кожній колонці та будувати об'єкти `RawEntry`, зчитуючи катакану, англійські визначення та інші деталі, наведені в останній колонці. Обробка також включає видалення пробілів та інших зайвих символів. На виході отримано файл `onomatopoeia1.json` зі структурованим словником, як можна побачити на *рис. 4.7*.

```
"\u30e9\u30d6\u30d6": [
  {
    "english": "*la di da* *la la la*",
    "details": "SFX for skipping and humming a happy tune."
  }
],
"\u30e9\u30d6\u30e9\u30d6": [
  {
    "english": "*romantic* *love love*",
    "details": "SFX that literally derives from the Japanese pronunciation of \"love-love\" as \"rah-bu rah-bu.\""
```

*Рисунок 4.7. — Приклад структури файлу onomatopoeia1.json. Реєстрові слова подані у форматі кодування UNICODE через особливість форматування файлу формату *.json.*

3) Написано код на Python (файл `add_SFX.py`), що призначений для завантаження японських звуконаслідувань з JSON-файлу та збереження результату у SQLite-базу даних.

Скрипт здійснює за допомогою засобів базової бібліотеки `sqlite3` [73] підключення до бази даних `combined_dictionary.db`, створює таблицю `SFX_japanese_onomatopoeia` та імпортує дані з файлу `onomatopoeia1.json` до неї за такою структурою: `hiragana` — запис у хірагані, `katakana` — запис у катакані, `definition` — англійські визначення, `details` — додаткові пояснення. Приклад отриманої таблиці в структурі бази даних продемонстровано на *рис. 4.8*.

hiragana	katakana	definition	details
Filter	Filter	Filter	Filter
ららら	ラララ	*la di da* *la la la*	SFX for skipping and humming a happy ...
らぶらぶ	ラブラブ	*romantic* *love love*	SFX that literally derives from the ...
らくらく	ラクラク	*too easy*	SFX derived from the adjective "easy"...
らんっ	ランッ	*beam* *intensely*	SFX for sudden attention in one's ...
らんらん	ランラン	*la la*	SFX for skipping and humming a happy ...
らんらん	ランラン	*intense* *fiery*	SFX for an eye expression, for ...

Рисунок 4.8. — Приклад структури бази даних з внесеними словниковими статтями «Japanese Manga SFX» до перекладу.

- 4) Останнім етапом обробки бази даних японської ономаічної лексики є автоматичний переклад словникових визначень з англійської на українську (файл SFX_trans.py).

Аналогічно до обробки словникових статей зі Словника Грінченка в Розділі 3, доцільним було використання OpenAI API [64]. Кожне визначення перекладено з таблиць SFX_japanese_onomatopoeia та japanese_onomatopoeia (для даних, зібраних з «nihongoresources» аналогічно) та збережено в таблиці з аналогічною структурою SFX_translated в межах тієї самої бази даних combined_dictionary.db. Повний код представлено у зазначеному файлі.

Автоматичний переклад не може бути і не є фінальним варіантом роботи, адже всі визначення слід перевірити та відредагувати вручну в інтерфейсі бази даних через середовище DB Browser [72] (рис. 4.9). Автоматичний переклад слугує лише допоміжним інструментом для оптимізації та пришвидшення робочого процесу.

hiragana	katakana	definition
かっか / ぴりっと	カッカ / ピリット	Смакова реакція на щось гостре або пряне.
かっかっか	カッカカッ	Сміх персонажа, який звучить як карикатурний або зловісний сміх...
かっかっか	カッカカッ	Часто використовується для передачі звуку каблуків або крейди п...
かっきり	カッキリ	Точна кількість або момент.
かっきーん	カッキーン	Відчуття сильного напруження або психологічного замороження.
かっちり	カッチリ	Міцний, добре зроблений.

Рисунок 4.9. — Приклад відредагованих словникових статей зі Словника SFX у базі даних.

В результаті цих етапів роботи зі словників nihongoresources та SFX було дібрано 735 та 1868 словникових статей відповідно. Така кількість зумовлена тим, що в ручному режимі видалено ті слова першого словника, що наявні в другому (у зв'язку з більш розлогими поясненнями у словнику SFX), а також об'єднано в один рядок через символ «/» кілька реєстрових слів, що мали однакові визначення та схожі форми вираження.

4.2.4. Інтеграція транскрипції ромаджі до словникових статей

Оскільки в укладених таблицях відсутня транскрипція японських слів, ми вважаємо за потрібне додати її для наочності та кращого розуміння. Міжнародним стандартом є так звана транскрипція японської мови ромаджі для передавання звучання мови літерами латинського алфавіту (модель Гепберна [12]). Саме її буде реалізовано через її переважну однозначність та зрозумілість усіма, хто вивчає/знає японську мову, а також через легку програмну реалізацію — наявні готові бібліотеки для автоматичної транскрипції японського тексту через ромаджі.

Для реалізації цього завдання створено скрипт у файлі `convert_to_romaji.py` [Додаток 1]. Він автоматично додає транскрипцію японських онomatопей (з хірагани) до бази даних `combined_dictionary.db` у вигляді ромаджі, а також генерує катакану з хірагани (оскільки для словникових статей з «Японсько-українського тематичного словника онomatопейчної лексики» відсутні також поля з катаканою. Їхня наявність бажана для більш гнучкого пошуку слів у базі даних).

Основне завдання коду — доповнити словникову БД новими полями для транскрипції. Використано такі бібліотеки Python: `sqlite3` [73] (взаємодія з локальною базою даних SQLite), `fugashi` [52] (морфологічний аналізатор для японського тексту), `rukakasi` [67] (конвертація з японських абеток у ромаджі за моделлю Гепберна).

Функція `kana_to_romaji(text)` — аналіз тексту за допомогою `fugashi` (розділяє на слова), отримання хірагани (`.feature.kana`) кожного слова, передавання у `rukakasi` для конвертації в ромаджі.

Приклад: "びっくり" → "bikkuri", "ドン" → "don".

Функція `hira_to_kata(text)` перетворення хіраґани в катакану: для кожного символу хіраґани (U+3041 — U+3096) додає +0x60 до юнікод-значення (це дає відповідний символ катакани). Приклад: "びっくり" → "ビックリ".

В результаті до таблиць `jp_onomatoroeia`, `SFX_japanese_onomatoroeia`, `japanese_onomatoroeia` додано поле `romaji` та поле `katakana` до таблиці `jp_onomatoroeia`. Результат збережено у базі даних `combined_dictionary.db`.

4.3. Лінгвістичний аналіз бази даних японської ономатопеїчної лексики

Загалом з трьох розглянутих джерел було дібрано 3399 унікальних словникових статей з урахуванням об'єднаних в межах однієї статті слів, що є варіантами (мають однакове визначення та не суттєво відрізняються за формою вираження).

Що стосується питання класифікації японських ономатопів, у межах нашої бази даних було вирішено на основі зібраних дефініцій приписати їм ті ж категорії, що представлені в базі даних української ономатопеїчної лексики. Така структура дозволить на наступному етапі роботи провести об'єднання двох таблиць за категоріями та легко дібрати відповідники в межах кожної окремо взятої категорії.

Водночас такий підхід не суперечить принципам класифікації японської ономатопеїчної лексики, розглянутий у п. 2.4. Розділу 2 (дослідження О. Кобелянської [15]): класи гісей-го та гіон-го будуть в центрі нашої уваги через наявність еквівалентів для цих класів в українській мові, що буде продемонстровано практично в Розділі 5. Ці два класи представлені широко в нашій базі даних та саме для них були призначені семантичні категорії звуконаслідувань, розглянуті в п. 3.5. Розділу 3.

Було покласифіковано 1537 словникових статей відповідно до зазначених 36 семантичних підгруп, а ще 75 одиницям була призначена категорія «вигук». 1052 одиниці не були покласифіковані через їхню належність до класу гітай-го — так званих міметичних слів, що не мають звуконаслідувальних еквівалентів в українській лексиці та відповідно відсутні в нашій базі даних. Пошук

відповідників до цих слів потребуватиме поглибленого дослідження, аналізу інших частин мови, таких як прислівники, прикметники, дієслова та застосування перекладацьких модуляцій. Пошуки вирішення цієї проблеми можуть послугувати мотивацією для проведення подальших наукових досліджень. Приклади таких слів з нашої бази даних:

- ***mura / mura-mura*** — *Несподіване внутрішнє бажання щось зробити.*
- ***meki-meki*** — *Помітно покращуватися, розвиватися на очах, усе більше і більше, краще і краще (стан помітного прогресу та розвитку).*
- ***meso-meso*** — *Киснути, хандрити; «опускати руки» (про стан людини, що втратила інтерес до життя чи засмутилась після якоїсь невдачі, нещастя, лиха тощо).*
- ***mechakucha*** — *Повний безлад або хаос, коли все навколо розкидане та неприбране.*

Такі одиниці класифікуються японськими мовознавцями як ониматопи насамперед через їхню форму: переважно двоскладова редуплікована структура, характерна для всіх звуконаслідувальних слів японської мови.

Натомість японські вигуки переважно не мають такої структури, однак на практиці з ними маємо аналогічну ситуацію, що склалася в українському мовознавстві (часте нерозрізнення вигуків та звуконаслідувальних слів). Саме тому виявлені вигуки були нами визначені та включені до класифікації аналогічним чином, як у базі даних українських ониматопів.

Що стосується розподілу за джерелами:

- 1) З «Японсько-українського тематичного словника ониматопейчної лексики» було дібрано 796 словникових статей.
- 2) Зі словника nihongoresources — 735 одиниць.
- 3) Зі словника SFX — 1868 одиниць.

Повна таблиця покласифікованих словникових статей доступна за посиланням в Додатку 3.

Висновки до Розділу 4

У результаті проведеної роботи було проаналізовано вміст трьох джерел для добору матеріалів поповнення бази даних японської ономатопейчної лексики, серед яких: «Японсько-український тематичний словник ономатопейчної лексики» [37], база даних ресурсу «nihongoresources» [61], користувацький електронний словник «Japanese Manga SFX» [57].

Було використано засоби мови програмування Python [66] для автоматизованої обробки вмісту вказаних джерел та їхнього внесення до реляційної бази даних. Використано такі програмні бібліотеки: sqlite3 [73], re [68], csv [46], itertools [56], dataclasses [47], typing [75], collections [43], json [58], fugashi [52], pykakasi [67].

Для автоматичного перекладу визначень з баз даних словників nihongoresources та «Japanese Manga SFX» з англійської на українську застосовано програмні засоби OpenAI API [64], після чого в інтерфейсі програми DB Browser [72] проведено редагування вручну.

Усі здобутки розділу збережені у базі даних файлу combined_dictionary.db [Додаток 2] та об'єднані в таблиці merged_onomatopoeia, де словникові статті представлені за такими полями: hiragana, katakana, romaji (реєстрове слово у трьох варіантах запису), definition (визначення), examples, examples_translation (приклади вживання та їхні переклади українською), similar (близькі за значенням ономатопи), source (одне з вказаних вище трьох джерел), category (приписана семантична категорія).

Проведено лінгвістичний аналіз бази даних, реєстр прокласифіковано на вигуки та звуконаслідування, у межах звуконаслідувань виокремлено 5 груп, що поділяються на 36 підгруп. Лексика класу гітай-го не була прокласифікована у зв'язку з відсутністю прямих еквівалентів у відповідній базі даних українських ономатопів, однак не була виключена з бази даних.

Загальна кількість відібраних слів з трьох різних джерел становить 3399 унікальних одиниць без урахування варіантів в межах однієї статті.

РОЗДІЛ 5. ПОБУДОВА СИСТЕМИ ПОШУКУ УКРАЇНСЬКИХ ВІДПОВІДНИКІВ ДЛЯ ЯПОНСЬКИХ ОНОМАТОПІВ

5.1. Обрахування семантичної подібності

Зібравши українські та японські ономатопи в межах структурованої бази даних, ми маємо розглянути два останніх завдання в межах цієї роботи: 1) поєднання одиниць двох мов шляхом автоматичного пошуку еквівалентів для реалізації функціоналу перекладацького словника; 2) побудова користувацького інтерфейсу електронного словника.

На противагу ручному формуванню словникової бази еквівалентів, ми пропонуємо автоматичний формат виконання цього завдання, основним підґрунтям для чого слугуватимуть приписані словниковим статтям на попередніх етапах семантичні категорії. Наш подальший алгоритм дій в межах цього пункту:

- 1) Векторизація словникових визначень (word embedding);
- 2) Обрахування косинусної подібності між векторами українських та японських визначень в межах кожної визначеної семантичної категорії;
- 3) Запис даних про семантичну подібність в окрему таблицю бази даних.

Word embedding («вкладання» слів, або векторизація слів) — один з методів сучасного NLP (Natural Language Processing) моделювання, процес представлення тексту у вигляді числового вектору, який можна обробляти математично [55]. Цей метод дозволяє автоматично оцінити схожість слів / речень відповідно до їхньої «близькості» у векторизованому семантичному просторі.

Косинусна подібність — це метрика, яка використовується для вимірювання схожості між двома векторами незалежно від їхньої довжини. У контексті NLP вона дозволяє оцінити семантичну близькість слів, фраз або речень, які вже були перетворені у вектори за допомогою певних програмних засобів [45]. Вона дозволяє обрахувати подібність між двома текстовими одиницями в діапазоні від -1 до 1 (від протилежних значень до ідентичності).

Однією з програмних бібліотек мови Python для виконання цих завдань є SentenceTransformer, у якій доступна велика кількість попередньо навчених нейромоделей, оптимізованих для задач векторизації, семантичної подібності, парафразування, кластеризації, пошуку тощо [70]. Конкретна модель, яку ми пропонуємо використати для виконання поставленого завдання є paraphrase-multilingual-MiniLM-L12-v2, що має такі особливості: мультилінгвальність (навчена на більше ніж 120 млн. параметрів та 100 мовах, у тому числі українській, що сприяє її використанню як для міжмовного семантичного порівняння, так і для монологічного), виявляє достатньо високу точність та є швидшою за інші (такі як LaBSE або distiluse-base-multilingual-cased-v2 [54]), що також є перевагою.

У цьому контексті наявність заздалегідь прописаних семантичних категорій в обох таблицях бази даних (українській та японській) насамперед сприяє більш точним результатам автоматичного визначення подібності, дозволяє програмі сконцентруватися на визначенні подібності між словами, що вже мають певну визначену схожість на етапі лінгвістичного аналізу, будучи віднесеними до однієї семантичної категорії.

Таким чином, нами було створено код мовою програмування Python, що реалізує семантичне порівняння визначень до українських та японських ониматопей у відповідних категоріях, використовуючи модель SentenceTransformer для обчислення косинусної подібності між векторними представленнями визначень та внесення відповідних даних до таблиці SQLite (файл semantic_similarity.py [Додаток 1]).

Використані програмні бібліотеки: sqlite3 [73] (взаємодія з базою даних SQLite), pandas [64] (опрацювання табличних даних), SentenceTransformer (отримання даних про семантичну подібність), tqdm [74], re [68] (розбиття рядків на частини через регулярні вирази).

Опис структури коду:

1) Завантаження моделі SentenceTransformer('paraphrase-multilingual-MiniLM-L12-v2');

2) Отримання спільних категорій для українських та японських отоматопів з бази даних `combined_dictionary.db` (таблиці `merged_ukrainian_onomatopoeia` та `merged_onomatopoeia`);

3) Створення таблиці результатів `semantic_similarity` в `combined_dictionary.db` за структурою: `ukr_word`, `ukr_definition`, `jp_definition`, `category`, `similarity` (вирішено взяти до порівняння реєстрове українське слово + українське визначення, що потенційно може посприяти точності виконання завдання);

4) Обчислення подібності: отримуються вектори за допомогою `model.encode()` та косинусна подібність за допомогою функції `util.pytorch_cos_sim()`; Збереження даних до таблиці з попереднього етапу.

5) Створення кеш-таблиці `jp_embeddings`, що використовується для попереднього зберігання ембедингів для прискорення майбутніх запитів, щоб модель надалі не обчислювала ті самі дані кожного разу. Структура таблиці: `query_form` (раніше заданий текст для збереження), `embedding` (вектор у форматі байтового об'єкта BLOB).

6) Збереження результатів.

Здобутки цього етапу роботи слугуватимуть підґрунтям для подальшої побудови функції пошуку українських відповідників для японських отоматопів.

Варто також зазначити, що при доповненні/редагуванні бази даних в майбутньому, достатньо буде запустити код з файлу `semantic_similarity.py` для переобрахунку семантичних подібностей, що за передбаченим алгоритмом не вплине на функціонал подальших етапів. Таким чином, відкривається шлях для розвитку й удосконалення бази даних отоматопів без втручання в програмний код.

5.2. Побудова функцій пошуку в базі даних

В межах електронного словника ми маємо за мету створити такий функціонал: 1) Пошук за базою даних словникових статей як японських, так і українських отоматопів, у тому числі за вказаною користувачем категорією. Передбачити виведення результатів найбільш схожих одиниць з бази даних у

разі неточного введення; 2) Пошук українських відповідників для японських ономатопів за методами, описаними у попередньому пункті.

Пошук семантично схожих японських відповідників для введених користувачем українських ономатопів ми не ставимо за мету з двох причин: 1) Кількісна асиметричність двох таблиць бази даних (кількість японських реєстрових слів значно більша, ніж українських у межах кожної категорії, що не сприяє точності роботи алгоритма); 2) Оптимізація швидкості роботи програми (обрахування семантичної подібності є відносно часово- і ресурсозатратним програмним алгоритмом); Таким чином, основне спрямування нашого електронного словника — переклад з японської на українську. Однак у разі подальшого розвитку реєстру словника зворотний функціонал може бути передбачений у майбутньому.

Для реалізації поставлених завдань нами було створено два скрипти мовою Python: `find_similar.py` — функція пошуку за таблицею японських ономатопів за вказаним словом та/або категорією та виведення найближчих українських відповідників; `find_ukr.py` — функція пошуку за таблицею українських ономатопів за вказаним словом та/або категорією. Програмний код доступний в Додатку 2. Розглянемо докладно кожен зі скриптів:

- 1) Код `find_similar.py` реалізує функцію `find_ukr_matches_for_jp_input`, яка шукає відповідність між японським словом або фразою (`input`) та знаходить найбільш семантично подібні українські відповідники, використовуючи попередньо збережену базу SQLite, ембеддинги та модель `SentenceTransformer`.

Використані бібліотеки: `sqlite3` [73] (Підключення до бази `combined_dictionary.db`), `pandas` [65] (опрацювання SQL-результатів у вигляді `DataFrame`), `re` [68] (регулярні вирази для розбиття рядків), `SentenceTransformer` [70] (Отримання та порівняння ембеддингів), `numpy` [63] (обробка масивів для зчитування BLOB із таблиці `jp_embeddings`).

Функція `find_ukr_matches_for_jp_input` має такі аргументи: `query_text` — японський запит (може бути `hiragana/katakana/romaji`); `category` — опціонально,

звужує пошук до певної категорії; `top_k` – кількість найкращих українських відповідників, що повертаються в результаті (за замовчуванням має значення 5). Передбачено такі моменти:

Якщо введено тільки категорію без слова, SQL-запит: отримує всі рядки з `merged_onomatoroeia` за вказаною категорією, повертається список таких результатів.

Якщо введено слово хіраганою, катаканою або транскрипцією ромаджі з або без категорії, виконуються такі етапи:

- кодування запиту у вектор `model.encode(query_text)`,
- отримання японських рядків з бази (таблиця `merged_onomatoroeia`),
- завантаження ембедингів з кешу `jp_embeddings`,
- за векторним форматом запиту формується список відповідників з японської таблиці, що за формою найбільш схожі до запиту (використовується `cosine similarity`),
- сортування кандидатів за значенням подібності,
- визначається, в якому з полів знайдено відповідник (`hiragana`, `katakana`, `romaji`),
- формування списку синонімів: інших варіантів написання цього ж слова (зазначені в базі даних через символ «/»),
- пошук відповідностей в українській таблиці `semantic_similarity`, за полями: `category` (спільна з японською), `jp_definition`,
- вибір п'яти найбільш схожих українських відповідників, сортування за `similarity` у порядку спадання,
- виведення результатів у форматі словника `json` (рис. 5.1.) з елементами `query`, `category`, `results` (містить всі поля таблиці бази даних для кожного знайденого слова та по 5 українських відповідників).

```
155 print(find_ukr_matches_for_jp_input('nyaa', 'Звіпи'))
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SQL HISTORY TASK MONITOR

```
{'query': 'nyaa', 'category': 'Звіпи', 'results': [{'match': {'hiragana': 'にゃあ', 'katakana': 'ニャア', 'romaji': 'nyaa'}, 'synonyms': [], 'definition': 'Нявкання або мурчання kota.', 'category': 'Звіпи', 'examples': None, 'examples_translation': None, 'source': 'Japanese Manga SFX spreadsh eet.', 'ukrainian_matches': [{'word': 'Курняв', 'definition': 'Звуконаслідування котячого нявчання .', 'similarity': 0.844}, {'word': 'Дряп / Драп / Драпць', 'definition': 'Звуконаслідування на поз начення дряпання, подряпини, наприклад пазурами kota.', 'similarity': 0.79}, {'word': 'Рняв', 'def inition': 'Звуконаслідування, що позначає позіхання kota.', 'similarity': 0.789}, {'word': 'Няв', 'definition': 'Звуконаслідування, що відтворює котяче нявкання.', 'similarity': 0.776}, {'word': '}
```

Рисунок 5.1. — Приклад виводу функції `find_ukr_matches_for_jp_input()` у файлі `find_similar.py` [Додаток 2].

Такий формат є необхідним для опрацювання на етапі побудови користувацького інтерфейсу електронного словника.

- 2) Код `find_ukr.py` реалізує функцію `find_ukrainian_words`, яка здійснює пошук українських звуконаслідувальних слів з бази даних `combined_dictionary.db`, відповідно до: введеного слова (`query_text`); та/або категорії (`category`); повертає результати у вигляді словників json.

Використані бібліотеки: `sqlite3` [73] (підключення до бази `combined_dictionary.db`), `pandas` [65] (завантаження таблиці з бази у `DataFrame` і фільтрація), `re` [68] (регулярні вирази для розбиття рядків).

Функція `find_ukrainian_words()` має такі аргументи: `query_text` – український запит (слово або його частина, за якою виконується пошук); `category` – опціонально.

Передбачено такі моменти:

- підключення до бази (`sqlite3.connect("combined_dictionary.db")`);
- попередня обробка аргументів (уніфікація до нижнього регістру, обрізання пробілів);
- завантаження всіх рядків з таблиці `merged_ukrainian_onomatopoeia`, отримуємо весь словник у `pandas.DataFrame` ;
- Фільтрація слова і/або категорії (рядки типу "дзень/дзеньк" розбиваються на окремі частини);
- Обробка випадку, коли результатів немає (Повертається об'єкт з порожнім

results і повідомленням "Не знайдено відповідників");

- Сортування результатів за word та формування вихідної структури (за полями таблиці бази даних);
- виведення результатів у форматі словника json (рис. 5.2.) з елементами query, category, results (містить всі поля таблиці бази даних для кожного знайденого слова).

```
{'query': 'няв', 'category': 'звірі', 'results': [{'index': 1, 'word': 'Курняв', 'definition': 'Звуконаслідування котячого нявчання.', 'category': 'Звірі', 'examples': None, 'source': 'Словник української мови за ред. Б. Грінченка.'}, {'index': 2, 'word': 'Няв', 'definition': 'Звуконаслідування, що відтворює котяче нявчання.', 'category': 'Звірі', 'examples': "1) Коли що було надума [кіт], то зараз до своєї хазяйки і озветься: “Няв, няв!” (Г. Квітка-Основ'яненко)\n2) Сміливе кошеня було нечемне, мабуть, бо, не сказавши навіть “няв”, втекло чимдуж до мами (О. Кониський)\n3) [Голоси звідусіль. (Дехто вдає звірячий та пташиний крик): ] Куку! Кукуріку! Гав-гав! Няв-няв! А кшшш! (Леся Українка)\n4) Ледь Василько до хати, як Вірка тут як тут, кішкою зазирає і няв-няв: – Ой-ой-ой, як ти поліпшав (В. Лис)\n5) Палюга брязнула по бантині, почулося голосне, злякане “няв”, і мото рошні зелені жарини щезли (Ю. Мушкетик)", 'source': 'Словник української мови у 20 томах.'}, {'index': 3, 'word': 'Рняв', 'definition': 'Звуконаслідування, що позначає позіхання kota.', 'category': 'Звірі', 'examples': None, 'source': 'З напрацювань видавництва «Nasha Idea».'}]}
```

Рисунок 5.1. — Приклад виводу функції *find_ukrainian_words()* у файлі *find_ukr.py* [Додаток 2].

Створивши функції з основним функціоналом нашого електронного словника, лишається лише розглянути хід роботи для побудови користувацького інтерфейсу для взаємодії з нашим проектом.

5.3. Побудова користувацького інтерфейсу електронного словника

Інтерфейсу взаємодії користувача з електронним словником онوماتопеїчної лексики ми пропонуємо створити у форматі вебсайту. Для цього, крім розглянутих вище функцій, нам знадобиться 1) файл з мовою розмітки HTML (задає структуру вебсторінки та визначає ієрархію її елементів) [53] у поєднанні з CSS (мова стилів, яка визначає зовнішній вигляд HTML-елементів) [48] та JavaScript (мова програмування, яка додає логіку та інтерактивність до сайтів, створює функціонал) [49]; 2) Серверний скрипт вебпрограми, реалізований з використанням Flask [51] — ядро вебсловника, зв'язує інтерфейс користувача із функціями пошуку (*find_ukr.py* і *find_similar.py*) та дозволяє запускати проєкт в онлайн режимі.

У фінальному варіанті наш проєкт матиме структуру, як

продемонстровано на *рис. 5.3*.

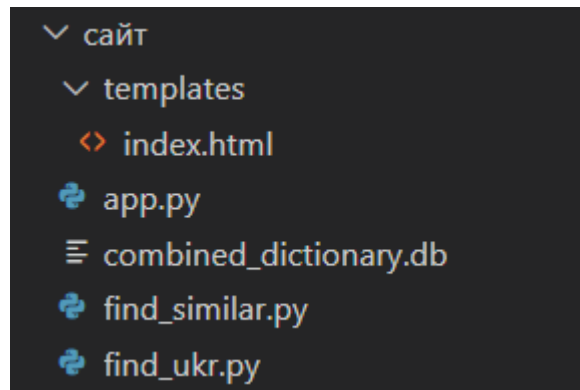


Рисунок 5.3. — Програмна структура електронного японсько-українського словника ономатопеїчної лексики [Додаток 2].

Де функції `find_similar.py`, `find_ukr.py` нами були розглянуті раніше, `combined_dictionary.db` — база даних, укладена нами в Розділі 3 та Розділі 4, файл `app.py` — серверний скрипт з використанням Flask [51], `index.html` — файл з мовою розмітки HTML (структура сайту).

Насамперед розглянемо структуру та функції файлу `index.html`.

Цей файл представляє повноцінну вебсторінку інтерфейсу електронного японсько-українського словника ономатопеїчної лексики, яка приймає запит користувача, дозволяє обрати мову та категорію пошуку, надсилає запит на сервер (`/search`), динамічно відображає результати пошуку (включно з синонімами, перекладом, прикладами, джерелом тощо).

Основна структура:

- `<head>` — містить заголовок та стилізацію сторінки (вбудований CSS відповідає за оформлення сторінки — фон, шрифт, стилі для кнопок, блоків результатів);
- `<body>` — інтерфейс користувача, містить поле вводу запиту (`<input>`), вибір мови пошуку (`<select id="lang" onchange="toggleCategory()">`), вибір категорії (`<select id="category">...</select>`), кнопку пошуку (`<button onclick="performSearch()">Пошук</button>`), порожній контейнер для результатів (`<div id="results"></div>`);
- JavaScript — містить основну функцію `performSearch()` для зчитування

полів query, lang, category (вибір здійснюється зі списку, зазначеного в частині <body>), надсилає асинхронний запит (fetch) на сервер app.py для отримання результатів виконання функцій find_similar.py та find_ukr.py згідно з вводом користувача, отримує словник json з відповідною структурою, рендерить кожен результат у поле results: слово, синоніми, визначення, категорія, приклади, переклад, джерело, українські відповідники + значення подібності.

Якщо нічого не знайдено:

```
container.innerHTML = "Нічого не знайдено.";
```

У разі помилки:

```
container.innerHTML = "Виникла помилка при запиті.";
```

- Зноска сайту (<footer>) — містить інформацію про авторство («Цей сайт створено в рамках дипломного проєкту. Автор: Ларіонов Олексій, 2025»), посилання на повний реєстр словника у GoogleSheets та на форму GoogleForms (рис. 5.4.) для пропозицій поповнення реєстру новими словами [Додаток 3].

Пропозиції для доповнення електронного японсько-українського словника онomatopēichної лексики

Автор: Ларіонов Олексій

larionovaleksej45@gmail.com [Switch accounts](#)

Not shared

* Indicates required question

Слово: *

Your answer

Визначення, короткий опис значення: *

Your answer

Приклади використання (якщо наявні):

Your answer

Submit Clear form

Рисунок 5.4. — Форма зворотного зв'язку з пропозиціями для доповнення реєстру словника.

Побудований файлом index.html користувацький інтерфейс представлено на *рис. 5.5*.

Рисунок 5.5. — Демонстрація інтерфейсу сторінки словника.

Тепер розглянемо структуру та функції файлу app.py [Додаток 2].

Це головний серверний скрипт вебпрограми, реалізований з використанням Flask [51] — вебфреймворку на Python [66]. Його роль — зв'язати інтерфейс користувача (index.html) із функціями пошуку (find_ukr.py і find_similar.py) та повертати результати у форматі json для відображення у браузері.

Основна структура:

- Імпорт з бібліотеки Flask необхідних модулів Flask (створення вебдодатку), request (доступ до параметрів запиту), jsonify (повернення даних у форматі json), render_template (рендер HTML-файлу index.html) та імпорт зазначених створених нами функцій.
- Ініціалізація додатку app = Flask(__name__) — створюється об'єкт app для запуску Flask-сервера.
- маршрут «/» (до головної сторінки сайту) — при відкритті вебсторінки в браузері рендериться HTML-шаблон index.html.
- Маршрут «/search» (основна функція) — приймає запити з JavaScript (через fetch у index.html), зчитує параметри query — введене слово, lang — мова запиту («jp» або «ukr»), category — вибрана категорія. Якщо «ukr» — викликається find_ukrainian_words з відповідним фільтром, якщо «jp» — викликається find_ukr_matches_for_jp_input з семантичним порівнянням.

- Повернення результату: `return jsonify(result)`.
- Запуск сервера: `if __name__ == "__main__": app.run(debug=True)`.

При запуску файлу `app.py` ініціалізується увесь проєкт та надається змога відкрити його у готовому вигляді через браузер за локальним хостингом. Проєкт готовий до хостингу в інтернет-мережі, проте наразі запускається локально.

5.4. Демонстрація результатів

Користувач обирає мову введення: японську або українську (за замовчуванням стоїть японська). При введенні японського слова хіраганою, катаканою або транскрипцією ромаджі користувач отримує список усіх подібних знайдених слів, їхні визначення, приклади та переклади (за наявності), семантичну категорію, джерело словникової статті та 5 найбільш схожих українських відповідників (рис. 5.6). Більше прикладів роботи функції добору українських відповідників див. у Додатку 4. Також надається можливість обмежити зону пошуку за задалегідь вказаною семантичною категорією.

Введіть слово:

Мова пошуку: Японська

Категорія: -- оберіть категорію --

(Пошук відповідників займає до 3 секунд після натискання)

1. にゃあ / ニャア / nya

Визначення: Нявкання або мурчання kota.

Категорія: Звірі

Джерело: Japanese Manga SFX spreadsheet.

Українські відповідники:

- Курняв — Звуконаслідування котячого нявканню. (Подібність: 0.844)
- Дряп / Драп / Драпць — Звуконаслідування на позначення дряпання, подряпини, наприклад пазурами kota. (Подібність: 0.79)
- Рняв — Звуконаслідування, що позначає позіхання kota. (Подібність: 0.789)
- Няв — Звуконаслідування, що відтворює котяче нявкання. (Подібність: 0.776)
- Мур-мур / Мурк — 1) Звуконаслідування, що відтворює муркотіння kota та інших котятих. 2) Те саме, що "муркотіння". 3) Уживається як присудок за знач. "муркотіти". 4) Нічого не говорити, мовчати. 5) Нічим себе не виявляти. 6) Ніяк не реагувати на що-небудь. 7) Не розбиратися в чомусь, не знати зовсім, не вміти чогось. (Подібність: 0.714)

Рисунок 5.6. — Приклад першого результату для введеного японського слова (загальна кількість знайдених статей у цьому випадку — 6).

За відсутності введеного слова, але наявності вказаної категорії, користувач отримує всі слова з бази даних, яким приписана ця категорія (список семантичних категорій розглянутий нами в п.3.3. Розділу 3), але без обрахування подібності з українськими ономатопами (рис. 5.7).

Введіть слово:

Наприклад: ばっ або Туп

Мова пошуку: Японська

Категорія: Комахи

Пошук (Пошук відповідників займає до 3 секунд після натискання)

1. うじゃうじゃ / ウジャウジャ / ujauja

Визначення: Роєм (про комах), роїтися (про велику кількість дрібних істот).

Категорія: Комахи

Приклади: 桜の木に毛虫がうじゃうじゃしているのには、びっくりしました。

Переклад: Здивувався, побачивши стільки гусені на вишні.

Джерело: Японсько-український тематичний словник ономапопеїчної лексики, Х. Еґава, О. Кобелянська.

2. ころころ / コロコロ / korokoro

Визначення: Цвірінь-цвірінь; цвіркотати, цвіркотіти; ха-ха-ха (про цвіркотіння комах; веселий сміх молодих дівчат тощо).

Категорія: Комахи

Приклади: 草の中で虫たちが、コロコロと鳴いている。もう秋だなあ。

Переклад: Цвірінь-цвірінь! – чути як цвіркочуть комахи у траві. Це вже осінь!

Джерело: Японсько-український тематичний словник ономапопеїчної лексики, Х. Еґава, О. Кобелянська.

Рисунок 5.7. — Приклад перших двох результатів для обраної категорії японських ономапонів (загальна кількість знайдених статей у цьому випадку — 12).

У разі відсутності введення та обрання жодної з категорії, результат — «Нічого не знайдено» (рис. 5.8).

Введіть слово:

Наприклад: ばっ або Туп

Мова пошуку: Японська

Категорія: -- оберіть категорію --

Пошук (Пошук відповідників займає до 3 секунд після натискання)

✗ Нічого не знайдено.

Рисунок 5.8. — Відсутність вводу.

У разі обрання введення українського слова користувач отримує аналогічно список усіх подібних знайдених слів, їхні визначення, приклади (за наявності), семантичну категорію, джерело словникової статті (рис. 5.9). За відсутності введення, але наявності вказаної категорії користувач аналогічно

отримує всі слова з бази даних, яким приписана ця категорія (рис. 5.10).

Введіть слово:

Мова пошуку:

Категорія:

(Пошук відповідників займає до 3 секунд після натискання)

1. Туп-туп / Тупць

Визначення: Звуконаслідування тупотіння: «топ-топ!».

Категорія: Рух

Приклади: Туп-туп ніженьками, дзень-дзень підківками.

Джерело: Словник української мови за ред. Б. Грінченка.

2. Тупу-тупу

Визначення: Звуконаслідування тупоту.

Категорія: Рух

Приклади: Тупу-тупу ногами, сколю тебе рогами.

Джерело: Словник української мови за ред. Б. Грінченка.

Рисунок 5.9. — Приклад перших двох результатів для введеного українського слова (загальна кількість знайдених статей у цьому випадку — 3).

Введіть слово:

Мова пошуку:

Категорія:

(Пошук відповідників займає до 3 секунд після натискання)

1. Абрюку

Визначення: Звуконаслідування, що виражає голубине воркування.

Категорія: Птахи

Джерело: Словник української мови за ред. Б. Грінченка.

2. Бруку

Визначення: Звуконаслідування, що виражає голубине воркування.

Категорія: Птахи

Приклади: Глянь, як голубка в парі брукає: бруку-бруку, бруку-бруку, бруку-ку!

Джерело: Словник української мови за ред. Б. Грінченка.

Рисунок 5.7. — Приклад перших двох результатів обраної категорії українських ономотопів (загальна кількість знайдених статей у цьому випадку — 54).

Висновки до Розділу 5

У межах цього розділу нашої роботи було розглянуто практичні аспекти побудови системи електронного словника за наявності попередньо створеної бази даних `combines_dictionary.db` [Додаток 2].

Нами було вирішене питання побудови функціоналів пошуку еквівалентів між японськими та українськими ономатопами шляхом векторизації словникових визначень та обрахування косинусної семантичної подібності між ними. Ця функція збережена у файлі `semantic_similarity.py` [Додаток 1] та дозволяє зберігати в окремій таблиці бази даних значення косинусної подібності. У разі поповнення або редагування бази даних новими реєстровими одиницями значення можуть бути легко переобчислені без порушення функціоналу решти програмного коду.

Було створено два скрипти мовою Python [66], що уможлиблює здійснення пошуку японських ономатопів та виведення українських відповідників (`find_similar.py`) та пошук українських ономатопів з бази даних (`find_ukr.py`).

Завдяки використанню мови розмітки вебсайтів HTML [53] у поєднанні з CSS [48] та JavaScript [49] було побудовано користувацький інтерфейс електронного словника (`index.html`), що реалізує весь зазначений функціонал. За допомогою вебфреймворку Flask [51] було об'єднано всі елементи проєкту та підготовано вебсайт для хостингу в мережі (`app.py`). Продемонстровано приклади взаємодії з системою.

Така структура проєкту дозволяє в подальшому розвивати реєстрову базу лексикографічної моделі. На створеному вебсайті додано посилання на форму для зворотного зв'язку користувача у разі бажання доповнити реєстр новими одиницями. Опубліковано повний зміст нашого словника у форматі електронної таблиці [Додаток 3]. Більше прикладів демонстрації функції добору українських еквівалентів для японських ономатопів представлено у [Додаток 4].

ВИСНОВКИ

У роботі було поставлено за мету створити електронний перекладацький словник з базою даних для ілюстрації одиниць звукозображувальної лексичної системи японської та української мов — ономатопеїчної лексики. В результаті проведеного дослідження мету було досягнуто: сформовано базу даних із 879 українських ономатопів (і вигуків) та 3349 японських, самостійно прокласифіковано їх за 36 семантичними категоріями (окремі категорії становлять вигуки та фразеологічно зв'язані вигуки, а також лексика гітай-го японської мови), створено користувацький інтерфейс словника у форматі вебсайту. Функціонал проєкту дозволяє здійснювати гнучкий пошук лексики двох мов, а також автоматично знаходити українські відповідники японським ономатомам. На нашу думку, цей ресурс є корисним для перекладачів, що працюють у сфері японсько-українського перекладу, зокрема перекладу японської манги.

У досягненні мети було виконано всі поставлені завдання як у межах теоретичної, так і практичної частини роботи, що дозволяє зробити такі висновки.

В українській мовознавчій традиції ономатопи були об'єктом розгляду небагатьох наукових досліджень, однак нам вдалося систематизувати ці знання, зокрема що стосується місця цієї лексики в системі мови, її класифікації як за структурно-формальним, так і семантичним принципом, визначити термінологічну базу, якою послуговуються дослідники при розгляді цього питання (ономатоп / звуконаслідування, вигук, звукозображувальна система мови). Основний масив теоретичної інформації стосовно ономатопеїчної лексики української мови, яким ми оперували на практиці, зібрано з праць таких дослідників, як І. Гаценко [8, 7, 5, 6, 4], Л. Мацько [17], О. Валігура [3]. Теоретична інформація послугувала підґрунтям для створення класифікаційної моделі нашої бази даних.

Було розглянуто аналогічне питання в межах японської мови, визначено відмінні та подібні риси в підходах до вивчення ономатопів в цих двох мовах.

Проаналізовано оригінальні праці японських мовознавців, таких як І. Таморі [30], К. Таджіма [29] та М. Танно [31], що дозволило укласти першоджерельне уявлення про класифікації, якими послуговуються в межах японської лінгвістики стосовно ониматопеїчної лексики. На основі теоретичної та практично отриманої інформації, можемо стверджувати, що ониматопеїчна лексика японської мови є значно більша та частотніша, порівняно з українською. Про це свідчить насамперед кількісна асиметрія укладеної бази даних двох мов та наявність широкого висвітлення цієї проблеми в межах японської лексикографії. У практичній частині роботи ми послуговувалися також здобутками теоретичних досліджень О. Кобелянської, зокрема цінним виявилось видання «Японсько-український тематичний словник ониматопеїчної лексики» [9]. Цей словник став одним з основних джерел для практичного укладання бази даних японських ониматопів, наряду з електронними джерелами «Japanese Manga SFX» [57] та «nihongoresources» [61].

Що стосується джерельної бази для укладання реєстру українських ониматопів, було опрацьовано 4 словники («Словник української мови у 20 томах» [38], «Словник української мови за ред. Б. Грінченка» [36], «Великий тлумачний словник сучасної української мови вид. «Перун» [32], «Великий електронний словник української мови» [33]), 1 електронне джерело (блог «livejournal» [59]) та спеціально надані в межах співпраці напрацювання українського видавництва японської манги «Nasha Idea» [60].

Проект може бути розвинений надалі в таких напрямках: розширення реєстру словника, покращення візуального наповнення інтерфейсу вебсайту, а також насамперед — хостинг сайту для загального доступу в мережі Інтернет, що потребує фінансових ресурсів.

Амбітною темою для майбутнім досліджень може стати створення перекладацького корпусу для ілюстрації модуляцій, що можливі при перекладі безеквівалентної японської ониматопеїчної лексики класу гітай-го.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Богдан Л., 2010. Структурно-семантичні особливості японської дитячої оноματοпоетичної лексики. Вісник КНУ імені Тараса Шевченка, серія «Східні мови та літератури», 16. с. 4–7.
2. Бондаренко І. П., Осадча Ю. В. упоряд., 2010. Японська література: Хрестоматія, том 1 (VII–XIII ст.). Київ: Видавничий дім Дмитра Бураго.
3. Валігура О. Р., Охріменко В. О., 2019. Специфіка сучасних підходів до вивчення ономотопей. Записки з ономастики, 22, с. 31–43.
4. Гаценко І.О., 2002. Синтаксична функція звуконаслідувальних слів української мови. Сіверянський літопис, 6, с. 97–100.
5. Гаценко І.О., 2003. Класифікація звуконаслідувальних слів української мови. Актуальні проблеми української лінгвістики: теорія і практика, 7, с. 122–127.
6. Гаценко І.О., 2003. Лексична характеристика звуконаслідувальних слів української мови. Наука і сучасність: Збірник наукових праць НПУ ім. М.П. Драгоманова, 37, с. 162–167.
7. Гаценко І.О., 2003. Типологічні особливості звуконаслідувальних слів. Кандидатська дис. НАН України.
8. Гаценко І. О., 2009. Типологічне зіставлення ономотопей у сучасному мовознавстві. Проблеми семантики, прагматики та когнітивної лінгвістики, 16, с. 129–135.
9. Егава Х. та Кобелянська О., 2017. Японсько-український тематичний словник ономотопейної лексики. Київ: Видавничий дім Дмитра Бураго.
10. Загнітко А. П., ред., 1998. Українська мова: словник-довідник. Донецьк: Центр підготовки абітурієнтів.
11. Звонська Л. Л., ред., Корольова Н. В., Лазер-Паньків О. В. та ін., 2017. Мора; Подовження. Енциклопедичний словник класичних мов, ВПЦ «Київський університет», с. 335.

12. Комаринська Т. К., Комісаров К. Ю., 2012. Теоретичний курс сучасної японської літературної мови, т. 1. Київ: Видавничий дім Дмитра Бураго.
13. Кобелянська О.І., 2016. Морфологічні, функціональні та звукосимволічні особливості ономотопеїчної лексики в японському поетичному тексті. Мова і культура, 18, т. 4, с. 334-344.
14. Кобелянська О., 2011. Вокальний звукосимволізм японської ономотопоеїчної лексики. XV Сходознавчі читання А. Кримського, с. 36–42.
15. Кобелянська О., 2013. Ономотопоеїчна система сучасної японської мови. Кандидатська дис. НАН України.
16. Коваленко О. О., 2012. Система японсько-української практичної транскрипції. Збірник наукових праць ННДІУВІ, 29, с. 197–215.
17. Мацько Л. І., 1985. Семантика вигуків та звуконаслідувань. Українська мова і література в школі, 2, с. 56–62.
18. Мельник О., 2011. Семантика вигука А в українській мові. Українське мовознавство, 42, с. 26–40.
19. Мосенкіс Ю.Л., 2005. Порівняльно-історичне і зіставно-типологічне мовознавство. Мова та історія, 84, с. 115–117.
20. Насікан З. С., 2015. Роль фоносемантичних засобів у створенні іконічності. Науковий вісник Міжнародного гуманітарного університету, 14, с. 186–188.
21. Охріменко В. О., 2015. Ономотопея як об'єкт вивчення фоносемантики. Наукові записки. Серія: філологічні науки, 138, с. 603–606.
22. Ржевська Д. О., 2023. Когнітивні особливості перекладу японської ономотопеїчної лексики. Вчені записки ТНУ імені В. І. Вернадського, т. 34, №2, ч. 1., с. 216–220.
23. Шистко А. О., 2023. Проблема вмотивованості слова. Міжнародна наукова конференція «Лексико-граматичні інновації», Дніпро, с. 174–176.
24. Шуст Л., 2014. Вигуки та вигуківі фразеологізми у творчості Т. Шевченка. Наукові записки НУ «Острозька академія», 47, с. 184–187.

25. Chang, A. S., 1990. Japanese-English Onomatopoeic Dictionary (Classification Type). Tokyo: Thaisyukan-shoethen.
26. Hisao K., Ikuhiro T., Lawrence S., Leslie J. E., 1996. Dictionary of Iconic Expressions in Japanese. Berlin: New York: Mouton de Gruyter.
27. Hida Y., Asada H., 2002. Gendai Giongo Gitaigo Youhou Jiten. Japan: Tokyodou Shuppan.
28. Inose H., 2007. Translating Japanese Onomatopoeia and Mimetic Words. Translation and Research Project 1. Spain: Universitat Rovira i Virgili, p. 97–116.
29. 田嶋 香織、2006. オノマトペ(擬音語擬態語)について. 日本語教育論集 16号、p. 194–205.
30. 田守 育啓、1998. 日本語多様な音と様態の表現オノマトペ. 日本音響学会誌54巻3号、p. 215–222.
31. 丹野 真智俊、2005. オノマトペ(擬音語・擬態語)を考える—日本語音韻の心理学的研究. あいり出版.

ЕЛЕКТРОННІ ДЖЕРЕЛА

32. Бусел В., 2005. Великий тлумачний словник сучасної української мови. Вид. «Перун». URL: <https://archive.org/details/velykyislovnyk> (Дата звернення: 05.06.2025).
33. Великий електронний словник української мови (BECUM) – GitHub-репозитарій Brown dict_uk. URL: https://github.com/brown-uk/dict_uk (Дата звернення: 05.06.2025).
34. Генеральний регіонально анотований корпус української мови (ГРАК). URL: <https://uacorporus.org> (Дата звернення: 05.06.2025).
35. Горох – Словник української мови. URL: <https://goroh.pp.ua> (Дата звернення: 05.06.2025).
36. Грінченко Б., ред., 1909. Словарь української мови. URL: https://chtyvo.org.ua/authors/Hrinchenko/Slovar_ukrainskoi_movy/ (Дата звернення: 05.06.2025).

37. Егава Х., Кобелянська О., 2016. Японсько-український тематичний словник ономатопеїчної лексики. URL: https://chtyvo.org.ua/authors/Kobelianska_Olha/Yaponsko-ukrainskyi_tematychnyi_slovnyk_onomatopeichnoi_leksyky/ (Дата звернення: 05.06.2025).
38. Словник сучасної української мови в 20 томах – Український мовно-інформаційний фонд НАН України. URL: <https://sum20ua.com/?wordid=0&page=0> (Дата звернення: 05.06.2025).
39. Словник.уа – онлайн словник. URL: <https://slovnyk.ua/index.php> (Дата звернення: 05.06.2025).
40. Уемура Сусуму. Курс лекцій з історії японської мови. URL: <http://www.philology.kiev.ua/php/26/bibl.html> (Дата звернення: 05.06.2025).
41. Asyncio – Asynchronous I/O. URL: <https://docs.python.org/3/library/asyncio.html> (Дата звернення: 05.06.2025).
42. BeautifulSoup – Documentation. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (Дата звернення: 05.06.2025).
43. Collections module – Python documentation. URL: <https://docs.python.org/3/library/collections.html> (Дата звернення: 05.06.2025).
44. Convertio – Онлайн-конвертер форматів. URL: <https://convertio.co> (Дата звернення: 05.06.2025).
45. Cosine Similarity – ScienceDirect. URL: <https://www.sciencedirect.com/topics/computer-science/cosine-similarity#:~:text=Cosine%20similarity%20is%20a%20measure,of%20the%20angle%20between%20them> (Дата звернення: 05.06.2025).
46. CSV module – Python documentation. URL: <https://docs.python.org/3/library/csv.html> (Дата звернення: 05.06.2025).
47. Dataclasses module – Python documentation. URL: <https://docs.python.org/3/library/dataclasses.html> (Дата звернення: 05.06.2025).
48. Developer Mozilla – CSS documentation. URL: <https://developer.mozilla.org/en-US/docs/Web/CSS> (Дата звернення: 05.06.2025).

49. Developer Mozilla – JavaScript documentation. URL: <https://developer.mozilla.org/en-US/docs/Web/JavaScript> (Дата звернення: 05.06.2025).
50. Docs aiohttp – Python async framework. URL: <https://docs.aiohttp.org/en/stable/> (Дата звернення: 05.06.2025).
51. Flask – Documentation. URL: <https://flask.palletsprojects.com/en/stable/> (Дата звернення: 05.06.2025).
52. Fugashi – MeCab wrapper for Python. URL: <https://pypi.org/project/fugashi/> (Дата звернення: 05.06.2025).
53. HTML Living Standard – WHATWG. URL: <https://html.spec.whatwg.org> (Дата звернення: 05.06.2025).
54. Hugging Face – DistilUSE Multilingual model. URL: <https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v2> (Дата звернення: 05.06.2025).
55. IBM – Word Embeddings. URL: <https://www.ibm.com/think/topics/word-embeddings> (Дата звернення: 05.06.2025).
56. Itertools module – Python documentation. URL: <https://docs.python.org/3/library/itertools.html> (Дата звернення: 05.06.2025).
57. Japanese Manga SFX Spreadsheet. URL: <https://docs.google.com/spreadsheets/d/1iXFIVmSzukN1PYD1JakdqMmrNNlc7TtkutDKevuKM0c/edit?usp=sharing> (Дата звернення: 05.06.2025).
58. Json module – Python documentation. URL: <https://docs.python.org/3/library/json.html> (Дата звернення: 05.06.2025).
59. LiveJournal – Raphnana. Блог. URL: <https://raphnana.livejournal.com> (Дата звернення: 05.06.2025).
60. Nasha idea – видавництво. URL: <https://nashaidea.com> (Дата звернення: 05.06.2025).
61. NihongoResources. URL: <https://www.nihongoresources.com> (Дата звернення: 05.06.2025).

62. NihongoResources GitHub mirror. URL:
<https://github.com/Pomax/nihongoresources.com> (Дата звернення: 05.06.2025).
63. NumPy – Documentation. URL: <https://numpy.org/doc/> (Дата звернення: 05.06.2025).
64. OpenAI API Reference. URL:
<https://platform.openai.com/docs/api-reference/introduction> (Дата звернення: 05.06.2025).
65. Pandas – Documentation. URL: <https://pandas.pydata.org/docs/> (Дата звернення: 05.06.2025).
66. Python 3.11.4 – Release info. URL:
<https://www.python.org/downloads/release/python-3114/> (Дата звернення: 05.06.2025).
67. Pykakasi – Documentation. URL: <https://pykakasi.readthedocs.io/en/latest/>
(Дата звернення: 05.06.2025).
68. Re module – Python documentation. URL:
<https://docs.python.org/3/library/re.html> (Дата звернення: 05.06.2025).
69. Requests – HTTP for humans. URL: <https://requests.readthedocs.io/en/latest/>
(Дата звернення: 05.06.2025).
70. SBERT – Sentence transformers. URL: <https://sbert.net> (Дата звернення: 05.06.2025).
71. SQLite – Official website. URL: <https://www.sqlite.org/index.html> (Дата звернення: 05.06.2025).
72. SQLite Browser – DB Browser for SQLite. URL: <https://sqlitebrowser.org>
(Дата звернення: 05.06.2025).
73. SQLite3 module – Python documentation. URL:
<https://docs.python.org/3/library/sqlite3.html> (Дата звернення: 05.06.2025).
74. TQDM – Progress bars in Python. URL: <https://tqdm.github.io> (Last accessed: 05.06.2025).
75. Typing module – Python documentation. URL:
<https://docs.python.org/3/library/typing.html> (Дата звернення: 05.06.2025).

ДОДАТКИ

Додаток 1. Повний програмний код для формування бази даних українських та японських ониматопів, включно з проміжними файлами.

Переглянути всі файли проекту та отриманих даних можна на GitHub репозиторії за посиланням:

<https://github.com/Lex0-228/DATABASE-CONSTRUCTION-Japanese-Ukrainian-Onomatopoeia-Dictionary>

Додаток 2. Повний програмний код японсько-українського електронного словника ониматопейчної лексики, готового до хостингу на сервері.

Переглянути всі файли проекту можна на GitHub репозиторії за посиланням:

<https://github.com/Lex0-228/Japanese-Ukrainian-Onomatopoeia-Dictionary>

Додаток 3. Посилання на вебсайті: альтернативний формат створеної бази даних та форма зворотного зв'язку.

Реєстр словника у форматі електронної таблиці GoogleSheets можна переглянути за посиланням:

<https://docs.google.com/spreadsheets/d/1ZWR51b7RxSXdSYVFvGzWCDwQvXLyGbS2x-2-70-HBxY/edit?usp=sharing>

Форма зворотного зв'язку GoogleForm, розміщена на створеному сайті для пропозицій доповнення реєстру:

<https://forms.gle/33NHU1i9iRkZXNYs8>

Додаток 4. Приклади роботи системи пошуку українських відповідників для японських онomatopie.

2. えっへ / エッへ / ehhe

Синоніми: ふ

Визначення: Звук короткого іронічного сміху — хе.

Категорія: Сміх

Джерело: Japanese Manga SFX spreadsheet.

Українські відповідники:

- Ха-ха — Звуконаслідування сміху. (Подібність: 0.904)
- Хе-хе — Звуконаслідування сміху. (Подібність: 0.9)
- Хихи — Звуконаслідування сміху. (Подібність: 0.859)
- Хи / Хи-хи — Звуконаслідування для вираження сміху. (Подібність: 0.857)
- Охохо — Звуконаслідування на позначення елегантного сміху. (Подібність: 0.853)

1. ひゅーひゅー / ヒューヒュー / hyuuhyuu

Визначення: Летіти зі свистом (про речі, що дуже швидко пролітають, розсікаючи повітря / кулі, швидкісні автомобілі, потяги тощо/).

Категорія: Рух

Приклади: 鉄砲の弾が、あっちからも、こっちからもひゅーひゅー飛んでくるんですから、生きて帰れたのが不思議なぐらいですよ。

Переклад: Звідусюди зі свистом летіли кулі, навіть не знаю, як я повернувся живим.

Джерело: Японсько-український тематичний словник онomatopieчної лексики, Х. Еґава, О. Кобелянська.

Українські відповідники:

- Фуш — Звуконаслідування, що описує швидкісний рух предмета, що пролітає повз. (Подібність: 0.726)
- Чкур — Звуконаслідування на позначення швидкої дії, "чкурнути": дуже швидко бігти, летіти і т. ін. (Подібність: 0.706)
- Юрк — Звуконаслідування на позначення свисту, швидкого руху, що розсікає повітря. (Подібність: 0.706)
- Вур — Звуконаслідування, що описує політ, кружляння. (Подібність: 0.683)
- Вууу — Звуконаслідування, що описує пролітання предмета повз, загрозу або напружену атмосферу. (Подібність: 0.681)

3. がたがた / ガタガタ / gatagata

Визначення: Бум-бум! Гучно вдаряться, громихати при хитанні (багаторазовий важкий гучний звук ударів при хитанні твердих предметів).

Категорія: Удар

Приклади: 地震のとき、急に地面がガタガタとゆれたので、びっくりしましたよ。

Переклад: Раптом земля з гуркотом затряслась через землетрус, що дуже мене налякав.

Джерело: Японсько-український тематичний словник онomatopieчної лексики, Х. Еґава, О. Кобелянська.

Українські відповідники:

- Торох — Звуконаслідування сильного удару, стуку. (Подібність: 0.884)
- Хряп — Звуконаслідування, що імітує різкі звуки ударів: бац, трах, шльоп. (Подібність: 0.857)
- Хвись / Хвиць — Звуконаслідування удару з розмаху по чомусь. (Подібність: 0.857)
- Хряпуть — Звуконаслідування, що імітує різкі звуки ударів: бац, трах, шльоп. (Подібність: 0.848)
- Тарах — Звуконаслідування удару, падіння, грохоту. (Подібність: 0.841)

1. くんくん / クンクン / kunkun

Визначення: Нюх-нюх! Сопіти, принохуватись, винюхувати; шморгати носом (звук, що утворюється носом при вдиханні та видиханні повітря;

Категорія: Ніс

Приклади: 1)子犬がクンクン鳴きながら近づいて来た。2)犬はいつもくくん何かの臭いをかいでいる。

Переклад: 1) Цуценя, принохуючись, наблизилось до мене. 2) Собаки завжди «нюх-нюх» – вишукують якийсь запах.

Джерело: Японсько-український тематичний словник ономотопеїчної лексики, Х. Еґава, О. Кобелянська.

Українські відповідники:

- Нюх-нюх — Звуконаслідування, що позначає дію "нюхати": вдихати через ніс який-небудь запах. (Подібність: 0.936)
- Шморг — Звуконаслідування, що описує утягувати носом повітря, особливо під час нежитю. (Подібність: 0.794)
- Чмих — Звуконаслідування фиркання. (Подібність: 0.657)

1. かきかき / カキカキ / kakikaki

Визначення: Звук царапання або швидкого писання чимось по паперу. «шкряб-шкряб».

Категорія: Дія рукою

Джерело: Japanese Manga SFX spreadsheet.

Українські відповідники:

- Жовк — Звуконаслідування на позначення копірсання, возіння з чимось. (Подібність: 0.72)
- Шпорт — Звуконаслідування на позначення копірсання, возіння з чимось. (Подібність: 0.698)
- Чик-чик — Звуконаслідування різання, від дієсл. «чикати». (Подібність: 0.694)
- Жмих — Звуконаслідування, що описує щось, що піддається вдавлюванню, стисканню. (Подібність: 0.684)
- Щип — Звуконаслідування на позначення щипання, затискання пальцями або дзьобом. (Подібність: 0.682)

1. うろうろ / ウロウロ / urouro

Визначення: Блукати, вештатися; кружляти довкіл, ходити навколо.

Категорія: Ходьба

Приклади: 1)たずねた家がわからなくて、30分もうろうろ歩いた。2)変な男がうちのまわりをうろうろしている。

Переклад: 1) Не знаючи, де будинок, блукав цілих 30 хвилин. 2) Дивакуватий чоловік вештається навколо будинку.

Джерело: Японсько-український тематичний словник ономотопеїчної лексики, Х. Еґава, О. Кобелянська.

Українські відповідники:

- Швєндь — Звуконаслідування на позначення дії "швєндяти": ходити туди-сюди (перев. без певної мети, без потреби), тинятися. (Подібність: 0.691)
- Крадь — Звуконаслідування на позначення дії "крастися": непомітно йти, пробиратися кудись, підходити до кого-, чого-небудь. (Подібність: 0.678)
- Човг / Човп / Чоп — Звуконаслідування шаркання ногами. (Подібність: 0.673)
- Дріб-дріб — Звуконаслідування на позначення дії "дріботіти": швидко перебираючи ногами, йти дрібними частими кроками, тьопати. (Подібність: 0.628)
- Чалап — Звуконаслідування на позначення повільної ходи по грязюці. (Подібність: 0.584)

1. げろげろ / ゲロゲロ / gerogero

Визначення: Кум-кум; квакати (звук, що видає велика жаба).

Категорія: Інші тварини

Приклади: 池からガマガエルのゲロゲロと鳴く声はずっと聞こえてくる。気になって、なかなか眠ることができない。

Переклад: Не можу заснути через безперервне квакання жаб на озері.

Джерело: Японсько-український тематичний словник ономотопеїчної лексики, Х. Еґава, О. Кобелянська.

Українські відповідники:

- Кум-кум — Звуконаслідування, що означає крик жаби. (Подібність: 0.903)
- Ква-ква — Звуконаслідування, що означає крик жаби. (Подібність: 0.892)
- Крумк — Звуконаслідування, що означає крик жаби, крука і т. ін. (Подібність: 0.849)