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

УКЛАДАННЯ СЛОВНИКА ОБСЦЕННОЇ ЛЕКСИКИ ТА СЛЕНГУ

Кваліфікаційна робота
на здобуття ОС «бакалавр»
студентки IV курсу
галузі знань 03 «Гуманітарні науки»,
спеціальність 035 «Філологія»
спеціалізації 035.10 «Прикладна
лінгвістика», ОПП «Прикладна
(комп'ютерна) лінгвістика
та англійська мова»
Поліни Володимирівни ШЕФЕР

Науковий керівник:
асистент кафедри української
мови та прикладної лінгвістики
Ярина ХОДАКІВСЬКА

Консультант:
кандидат техн. н., доц.
Микола КОСТІКОВ

«Допущено до захисту»
Протокол засідання кафедри
української мови та прикладної лінгвістики
від «5» 06 20__ року №16

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

КИЇВ – 2026

ЗМІСТ

АНОТАЦІЯ

У сучасному цифровому середовищі особливого значення набуває створення електронних лексикографічних ресурсів. Одним із недостатньо опрацьованих напрямів української лексикографії залишається опис обсценної лексики та сленгу, які активно функціонують у повсякденному мовленні, медійному просторі та соціальних мережах.

Метою кваліфікаційної роботи є укладання електронного словника українських обсценізмів та сленгу, створення бази даних для зберігання лексикографічного матеріалу та розроблення вебзастосунку для роботи зі словником. Для досягнення поставленої мети було опрацьовано наукову літературу з проблематики обсценної лексики, сленгу та електронної лексикографії, здійснено добір і систематизацію лексичного матеріалу, спроєктовано структуру бази даних та реалізовано вебінтерфейс словника.

У роботі використано матеріали словника Лесі Ставицької «Українська мова без табу. Словник нецензурної лексики та її відповідників. Обсценізми, евфемізми, сексуалізми», словника Катерини Бобровник та онлайн-словника Slang Zone. У результаті було відібрано 1790 лексем, які після опрацювання було згруповано у 1043 словникові статті. Для зберігання даних створено реляційну базу даних SQLite, а для забезпечення доступу до словника розроблено вебзастосунок мовою PHP.

Практичне значення роботи полягає у створенні функціонального електронного словника обсценізмів та сленгу, який може використовуватися як лексикографічний ресурс, а також як інструмент для прикладних досліджень з комп’ютерної лінгвістики. Результати роботи можуть бути використані для подальшого розвитку електронних словників української мови, аналізу мовлення соціальних мереж чи стилістично маркованої лексики.

Ключові слова: *комп'ютерна лексикографія, обсценна лексика, сленг, словник обсценізмів, база даних, SQLite, PHP, вебзастосунок, комп’ютерна лінгвістика.*

ABSTRACT

In the modern digital environment, the development of electronic lexicographic resources plays an increasingly important role in the systematization, storage, and analysis of linguistic data. One of the least explored areas of Ukrainian lexicography is the description of obscene vocabulary and slang, which are widely used in everyday communication, media discourse, and social networks.

The aim of this qualification paper is to compile an electronic dictionary of Ukrainian obscene vocabulary and slang, develop a database for storing lexicographic data, and create a web application for working with the dictionary. To achieve this goal, theoretical studies on obscene vocabulary, slang, and electronic lexicography were reviewed, lexical material was collected and systematized, a database structure was designed, and a web interface was implemented.

The study is based on the dictionary by Lesia Stavytska “Ukrainian Language without Taboo. Dictionary of Obscene Vocabulary and Its Equivalents. Obscenisms, Euphemisms, Sexualisms”, the dictionary by Kateryna Bobrovnyk, and the online dictionary Slang Zone. As a result, 1,790 lexical units were collected and organized into 1,043 dictionary entries. A relational SQLite database was created to store the data, and a PHP-based web application was developed to provide access to the dictionary.

The practical significance of the study lies in the creation of a functional electronic dictionary of obscene vocabulary and slang that can be used both as a lexicographic resource and as a tool for computational linguistic research. The results may serve as a basis for the further development of Ukrainian electronic dictionaries, studies of social media discourse, and analyses of stylistically marked vocabulary.

Keywords: *computational lexicography, obscene vocabulary, slang, dictionary of obscenisms, database, SQLite, PHP, web application, computational linguistics.*

ВСТУП

Лексичний склад будь-якої мови є складною та динамічною системою, яка постійно змінюється під впливом суспільних, культурних і технологічних чинників. Поряд із нормативною літературною лексикою в мовленні функціонують такі стилістично марковані мовні одиниці, як сленгізми, жаргонізми, просторічні слова та обсценізми, які тривалий час залишалися поза межами традиційної лексикографії або були представлені в ній лише частково.

Особливої актуальности такі лексичні одиниці набувають в контексті російсько-української війни, яка суттєво вплинула на лексичний склад української мови та спровокувала появу значної кількости мовних новотворів, зокрема й обсценного характеру. З'явились нові словотвірні моделі на базі вже наявних нецензурних коренів, виникли гібридні форми, що поєднують обсценні елементи з власними назвами, топонімами та військовою термінологією.

Метою дослідження є укладання електронного словника обсценної лексики та сленгу (на базі словника Катерини Бобровник, словника Лесі Ставицької та вебсторінки Slang Zone).

Для досягнення поставленої мети було визначено низку завдань:

- **розкрити поняття обсценної лексики** та охарактеризувати стан її вивчення в українській лінгвістиці, окресливши наявні підходи до визначення та класифікації цього явища. У межах цього завдання також передбачено опрацювання теоретичної літератури за темою дослідження з метою формування належної наукової бази;
- охарактеризувати чинники та **специфіку варіятивности обсценізмів**, з'ясувавши, які екстра- та інтралінгвістичні фактори зумовлюють їхню формальну та семантичну мінливість. Це дасть змогу глибше зрозуміти природу досліджуваних одиниць та механізми їхнього функціонування в мові;
- на основі матеріялів зі словника Катерини Бобровник **лематизувати обсценні словоформи** (паралельно планується опрацювання словника Лесі Ставицької

«Українська мова без табу: словник нецензурної лексики та її відповідників» та добір із нього відповідних лексем і дефініцій).

- **відбірати сучасні сленгізми із вебпорталу Slang Zone;**
- укласти бази даних словника та розробити для нього програмний інтерфейс за допомогою відповідних інструментів кодування.

Об'єктом дослідження є обсценна лексика української мови, зібрана наприкінці 20 – на початку 21 ст.

Предметом дослідження є лексикографування обсценної лексики та розроблення застосунку для словника.

Методи дослідження:

- описовий;
- лексикографічний (укладання словника);
- аналіз і синтез;
- емпіричний метод.

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків.

РОЗДІЛ 1.
ДОСЛІДЖЕННЯ ОБСЦЕННОЇ ЛЕКСИКИ В УКРАЇНСЬКОМУ МОВОЗНАВСТВІ. ВАРІЯТИВНІСТЬ ОБСЦЕНІЗМІВ

1.1. Стан вивчення обсценізмів в українському мовознавстві

В останні роки в лінгвістиці дедалі сильніші позиції займає дескриптивний підхід дослідження мовних явищ, зміщуючи прескриптивний. Ця тенденція свідчить про підвищення інтересу не стільки до мови як системи, скільки до мовлення як продукту мовної діяльности людини, індивідуальних мовних актів. На тлі поширення дескриптивізму науковці все частіше цікавляться проблемами, що раніше вважалися табуйованими або не вартими вивчення через свою невідповідність нормі. Однією з мало досліджених лексичних груп є обсценізми.

Термін *обсценний* походить від латинського слова *obscena*, що має кілька значень: 1) непристойні вчинки або слова, 2) статеві органи, 3) задня частина тіла, 4) виверження (Ставицька, 2008, с. 16). З часом це слово засвоїла французька мова як *obscéne*, звідки воно перейшло в англійську (*obscene*), а згодом – у термінологію українського мовознавства. У загальному вжитку поняття обсценности включає в себе як вербальні (лексичні) так і невербальні прояви (жестова мова тощо) відразливої поведінки. Обсценна лексика зокрема − часто вульгарні табуйовані слова та вислови, які мовці сприймають як непристойні.

Слово стає непристойним тоді, коли виражає референцію до певного образу, що викликає відчуття відрази. На думку Е. Берна, найсильніші непристойності стосуються запаху і смаку, а також слизького дотику (Berne, 1970).

Закордонні дослідження обсценної лексики охоплюють широкий спектр дисциплін — від лінгвістики та фонології до нейробіології. Так, наприклад, дослідників цікавить, чи мають лайливі слова універсальні звукові характеристики, які роблять їх придатними для вираження сильних емоцій. Шірі Лев-Арі та Раян Маккей у своєму дослідженні 2022 року виявили крос-

лінгвістичну закономірність: у лайці багатьох мов (іврит, хінді, угорська, корейська) спостерігається відсутність апроксимантів (сонорних звуків, таких як *l, r, w, y*). (Shiri Lev-Ari & Ryan McKay, 2022) Також вони довели, що «евфемістична лайка» (наприклад, англійське *darn* замість *damn*) часто створюється шляхом додавання саме цих «м'яких» звуків. (Shiri Lev-Ari & Ryan McKay, 2022)

Бенджамін К. Берген він зазначає, що англійські односкладові обсценізми частіше закінчуються на проривні приголосні (наприклад, *k, t*), що додає їм фонетичної різкости. (Bergen, 2016)

Іноземні дослідники також приділяють увагу типології обсценізмів у різних культурах. Магнус Л'юнг провів масштабне порівняльне дослідження англійської та ще 24 мов (Ljung, 2011). Він розробив типологію лайки та простежив її історію від Давнього Єгипту до наших днів (Ljung M, 2011). Ешлі Монтегю у своїх роботах з історії обсценної лексики виділив 14–15 категорій лайливих слів, включаючи звернення до надприродних сил, сакральних тем, назв тварин та соціяльних образ (Montagu, 1967).

У вітчизняному мовознавстві дослідження обсценізмів було основним науковим зацікавленням Лесі Олексіївни Ставицької. Працюючи над цією темою, вона уклала словники: «Український жарґон» (2010), «Українська мова без табу. Словник нецензурної лексики та її відповідників» (2008) та інші. А також написала монографію «Арґо, жарґон, сленґ: соціяльна диференціація української мови» (2005). Масова авдиторія сприйняла її роботи критично, вбачаючи у ній більше заохочення до вживання нецензурної лексики, ніж наукове дослідження. У передмові до подальших видань своїх словників науковиця пояснила, що її наміри були витлумачені хибно. За її словами, обсценізми підлягають вивченню на тій самій основі, на якій вивчаються будь-які інші соціолекти. Ставицька зазначає, що протягом багатьох років українські лінгвісти ігнорували нецензурну лексику, навішуючи на неї ярлики «деградації» та «духовної вбогости» (Ставицька, 2008, с. 7).

Відсутність належного дослідження обсценізмів вона пояснює не тільки «невротичними пуристичними тенденціями», поширеними по всьому світу (Ставицька, 2008, с. 11), а й потребою утвердження державного статусу української мови. Бажаючи зберегти еталонну мову, філологи-науковці часто закликають до її чистоти і виструнчування за єдиною нормою. Це призводить до того, що живе мовлення багатьох категорій мовців лишається поза полем зору науковців, оскільки узус неминуче відрізняється від прописаних правил. Усього, що знаходиться поза нормою, уникають як такого, що паплюжить мову. Крім звуження предмету лінгвістичних досліджень, надлишковий пуризм спричиняє дискомфорт для мовців, що починають ставитись до мови як до чогось сакрального і недоторканого. На думку Лесі Ставицької, "говорити українською замість радості мови, задоволення від мови часто зводиться до обов'язку перед мовою, психологічного дискомфорту та невдоволення собою і... мовою" (Ставицька, 2003, с. 12).

Леся Ставицька поділяє обсценізми на такі основні групи слів та їхні похідні:

1. Найменування «непристойних», соціяльно табуйованих частин тіла − «сороміцькі слова»: *хуй, пизда, мудо, срака.*
2. Найменування процесу здійснення статевого акту: *їбати.*
3. Назви повії, продажної жінки: *блядь.*
4. Найменування фізіологічних функцій (відправлень): *срати, сцяти*.
5. Назви «результатів» фізіологічних відправлень: *гівно*. (Ставицька, 2008, с. 18)

Ставицька вирішує термінологічну неоднозначність, уводячи у своїх дослідженнях такі поняття як *матизм* на позначення власне мату, а також *інвектива* – синонім до обсценної лексики. Цими термінами я також надалі буду послуговуватись.

Так, тільки перші три з наведених п'яти груп включають у себе лексику власне мату, пов'язану з сексуальною діяльністю людини. Таким чином, матизми входять у категорію обсценізмів, а обсценна лексика у свою чергу є поняттям

ширшим за матірну лексику. Усі матизми та їхні похідні можна назвати обсценізмами, проте не всі обсценізми є матами.

Стосовно класифікації власне українського обсценного прошарку, то він поділяється на такі групи лексем та похідні від них:

1) власне матизми − слова, пов'язані з сексуальною діяльністю: *їбати, пизда, хуй, манда, мудак, блядь*;

2) обсценна лексика сексуальної сфери: *залупа, секель; дрочити, мінет, гондон, менстра; поц, підарас, хер; курва, сука;*

3) обсценізми несексуальної сфери: *срати, сцяти, бздіти, дристати, пердіти; гівно, срака, жопа, задниця, гузно.* (Ставицька, 2008)

1.2. Комунікативно-прагматичні аспекти обсценного лексикону

За прагматичним аспектом обсценний лексикон поділяється на такі підтипи:

1. Посили, в основі яких лежить прагнення адресата відсторонитись від адресанта, відіславши його в певне місце «чужого» простору, тобто тілесного низу (*пішов на хуй, іди в сраку*). Такими посилами адресат виражає бажання не мати справи зі співрозмовником, не бачити його і не чути, і водночас прагне образити адресата.

2. Матірні вислови, що містять дієслово на позначення здійснення статевого акту і спрямовуються на матір адресата або на нього самого: *йоб твою мать; їбать мене в дишло*.

3. Відмови – в основі яких відмова адресата в бажанні адресанта, поєднана з образою: *хуй тобі в сраку.*

4. Індиферентиви − виражають байдужість, самовідсторонення, неохочу згоду: *єбись конем, хуй з тобою, до сраки карі очі.*

5. Пейоративи − це власне лайлива лексика, яка передає оцінку розумових чи фізичних здібностей людини, її характеру тощо. Часто вживається у поєднанні з прикметниками: *мудак, блядь, пизда, мандавошка, розпиздяй* абощо.

6. Божба − клятви, запевнення в правдивості висловленого: *бля(дь) буду, щоб хуя я ніколи не побачила.*

7. Прокльони: *Бодай тебе коні їбали! Напився б з гівна юшки!*

8. Вигуки і вставні слова: *йокелемене, бля(дь), йобані очі* (Ставицька, 2008, с. 26).

За функціональною характеристикою обсценізми можуть:

1) мати власне емотивну функцію і виражати обурення, гнів, здивування тощо;

2) вживатися мовцем безадресно з метою заповнення павз. У такому випадку з фонетичної точки зору в мовленні обсценізми набувають функції хезитаційних явищ, уподібнюючись до інших засобів вербальної (вокалізації, розтягнення звуків, слова паразити) та невербальної павзації (покашлювання, сміх, скрегіт зубів).

Серед обсценних хезитаційних явищ найуживанішими є *бля, блін* та похідні від *блядь.*

1.3. Особливості української обсценної лексики

Л. Ставицька аналізувала чимало компаративних досліджень обсценних прошарків різних мов світу. Як зауважує дослідниця, деякі науковці висували гіпотези, що велика кількість нецензурної лексики в мовах тих чи інших національностей пов’язана з їхньою войовничістю та агресивністю (араби, угорці, росіяни). Велике розмаїття обсценізмів також пов’язують з імперським характером народу, кліматом та місцем проживання. Спостереження продемонстрували, що чим ближче на південь, тим яскравіша лайка; мешканці морських держав найчастіше вдаються до міцного слова. Тим не менш достатньо науково обґрунтованого зв’язку між національним фактором та багатством обсценної лексики в мові досі не встановлено. (Ставицька, 2008)

У праці Лесі Ставицької, відповідно до національних ціннісних систем народів європейського ареалу, наведено поділ мов за такими типами обсценних стратегій:

1) «анально-екскремальний» (Shit (Scheiss)-культура): німецька, чеська, англійська, французька;

2) «сексуальний» (Sex-культура): російська, сербська, хорватська, болгарська;
3) «сакральний» (Sacrum-культура): романські мови, частково − чеська, словацька, польська.

Наведений поділ має умовний характер, оскільки різні стратегії співіснують у межах окремої лінгвокультури, проте переважання певного лексикону на тому чи іншому історичному етапі своєрідно віддзеркалює концептуальні засади буття кожного етносу, його стереотипи та ціннісну систему. Інвектива є табуйованою в будь-якій мові, отже, домінуючий тип обсценної стратегії може вказувати на те, які теми з-поміж усіх непристойних викликають найбільшу відразу у носіїв. Ставицька припускає, що якщо, наприклад, мові властива Shit-культура, то можна передбачити, що народ, котрий нею послуговується, є винятково чистим та охайним, оскільки екскрементальні лексеми і вислови викликають в мовців найбільш негативні асоціації.

Так, стверджує науковиця, у Західній Європі найохайнішими є німці, тому їхня лайка базується на скатологізмах – екскремальних лексемах. Не можна втім провести чітку межу між націями, спираючись на цю класифікацію, і віднести їх до виключно “порядних” чи “цнотливих”, оскільки зазвичай у мові співіснують декілька обсценних стратегій.

У словнику Лесі Ставицької наведено приблизне кількісне співвідношення обсценних сталих словосполучень в українській мові. Серед підрахованих словникових одиниць можна зафіксувати дві обсценні стратегії (Sex-культура та Shit-культура). Кількісна характеристика допомагає простежити домінанту: *їбати*, *їбатися* − до 20, *срати* − 46, *пизда* − 38, *хуй* − 57, *срака* – 151 , *жопа* – 76 (Ставицька, 2008). Можна стверджувати, що для української лайки більш властивим є скатологічний тип.

Фразеологізми відображають світогляд етносу повніше за його лексичне наповнення. Так, за Ставицькою, копрологічні преференції українців постають у фразеології: *своя пизда не пахнет − своє гівно не смердить* (неприємні якості

неможливо відрефлексувати), *как до пизды дверцы − до дупи дверцята* (ні туди ні сюди), в *пизде зубов нет − не шукай в сраці зубів*.

Леся Ставицька зауважує, що історично власне матизми як лайка не були властиві українській лінгвокультурі. В українському селі матірні слова не були в побутовому вжитку навіть у чоловіків; заборонялось лаятись в хаті, при дітях, жінках, старих людях та біля води.

В умовах глобалізації межі між обсценними культурами стають розмитими і взаємопроникними. Так, сучасна українська лайка сформована за значного впливу російських та польських обсценізмів.

Через свою експресивну функцію обсценна лексика нерідко стає складовою виступів медійних осіб та політичних діячів. Вона використовується з метою привернення уваги та підсилення емоційного впливу на адресата. Обсценізми вживають як в усних виступах і промовах, так і в письмових текстах в соціяльних мережах.

У ході навчання за освітньою програмою я виконала проєкт, присвячений аналізу мовлення українських політиків у соціяльних мережах. Одним із завдань проєкту було виявлення обсценної лексики в корпусі політичних текстів, укладеному студентами групи. Ця робота дозволяє продемонструвати можливості практичного використання укладеного «Словника обсценізмів та сленгу» для автоматизованого аналізу текстових даних.

Для виявлення обсценної лексики було використано платформу TextAttributor, а також «Словник обсценізмів та сленгу», укладений мною в межах попереднього курсового дослідження та надалі розширений під час підготовки кваліфікаційної роботи.

З метою автоматизації аналізу було розроблено програму мовою Python, яка здійснювала лематизацію текстів за допомогою бібліотеки Stanza. Після цього отримані леми порівнювалися зі списком лексем зі «Словника обсценізмів та сленгу». Такий підхід дозволив перевірити можливість використання словника не лише як довідкового лексикографічного ресурсу, а й як інструмента для комп’ютерного аналізу мовних даних.

Результати автоматичного аналізу засвідчили наявність обсценних лексем у текстах чотирьох із двадцяти чотирьох досліджених політиків. Для Мар'яни Безуглої та Олексія Гончаренка кількість знайдених обсценізмів становила 1, для Бориса Філатова – 3. Найбільшу кількість нецензурної лексики було виявлено в дописах Ірини Фаріон – 13 лексем. Водночас траплялися окремі випадки хибних спрацювань, коли програма помилково ідентифікувала нейтральні слова як обсценізми. Зокрема, у тексті Віталія Кличка програма визначила слово «переїхати» як обсценізм. Крім того, ручна перевірка текстів показала, що частина обсценної лексики залишилася невиявленою через відсутність деяких лексем у словнику або використання авторами графічно змінених («зацензурених») форм слів. Наприклад:

- в текстах Ірини Фаріон — «задниця», «тупорилість» (цих лексем бракує в словнику);
- в текстах Бориса Філатова — «н@піздюнкали», «бл@ть»;
- в текстах Олексія Гончаренка — «ПІЗДОВ», «піздец», «заєбали».

Програмі не вдалося відшукати ці лексеми, оскільки, по-перше я зосереджувалася переважно на більш маркованих обсценізмах, і, по-друге, я не включала до словника зацензуровані варіянти слів із використанням символів на кшталт «@», а також не передбачила подібні винятки у самому коді програми.

1.4. Мовна варіятивність у фокусі досліджень іноземних науковців

Соціяльність мови спричиняє її варіятивність. Ця варіятивність проявляється не тільки в паралельних граматичних і лексичних формах всередині мови. Існує чимало так званих плюрицентричних мов – таких, що мають окремі регіональні стандарти, що відрізняються цілою низкою правил та визнані офіційними в різних державах.

Ненормованість такого мовного пласту як обсценна лексика спричиняє виникнення широкого спектру варіянтів обсценізмів.

За визначенням Сепіра, «загальновідомим є те, що мова варіятивна» (Парфіненко, 2013). Мова варіюється на багатьох рівнях: фонетичному,

морфологічному, лексичному. З усіх перерахованих найпоширенішим є останній тип, хоча варіятивність достатньо представлена й на інших мовних рівнях. Так, в англійській мові співіснують лексичні варіянти *felloe – felly, infantile – infantine, mediatory – mediatorial*, фонетичні варіянти ['gærɑ:ʒ] – ['gæridʒ] – ['gærɑ:dʒ] для *garage* або [lef'tenənt] – [le'tenənt] для *lieutenant,* графічні варіянти *draught – draft, nosey – nosy, endue – indue* (Парфіненко, 2013). За Парфіненко, варіятивність є мовною надлишковістю, що призводить до подальшого розвитку мови (Парфіненко, 2013). Кілька варіянтних форм є конкуруючими. З плином часу одна з форм закріплюється в мові, стаючи нормативною, а інша забувається або ж набуває нового лексичного значення (як у морфологічних варіянтах historic/historical; economic/economical) (Парфіненко, 2013). Така конкуренція мовних засобів зазвичай призводить до одного з двох наслідків:

1. Одна з варіянтних форм переважає за статистичними показниками і таким чином стає загальновживаною, витісняючи решту можливих форм. Це природний процес, який часто відбувається в історії мови, і призводить до важливих лінгвістичних змін; саме частотність використання тих чи інших форм відіграє в ньому значну роль.

2. “Дублетна пара залишається стабільною, усі мовні форми входять в узус функційно рівноправно завдяки деякій різниці в лексичному чи граматичному значеннях. Так, уже в староанглійській мові співіснували парні форми "shined – shone" як перехідна та неперехідна форми. Деякі з середньоанглійських дублетів розвинули такі відмінності й залишились у мові до цього часу.” (Парфіненко, 2013).

Поява дублетів пояснюється соціолінгвістичним фактором: дублети виникають через вплив діялектів та внаслідок мовних контактів і конкурують у використанні доти, доки та чи інша форма не перемагає. Незважаючи на це, вони можуть стабільно співіснувати у мовній єдності, за умови диференціяції за значенням, перестаючи бути дублетами. Термін *дублети* не можна застосовувати відносно обсценізмів, оскільки для цього особливого лексичного пласту є властивою наявність значної кількости варіянтів, що нерідко перевищує дві

одиниці. На мою думку, “забуванню” менш уживаного дублету може сприяти уніфікація мови шляхом укладання словників, де фіксується нормативний варіянт. Так, оскільки обсценна лексика не є уніфікованою і за своєю функцією не потребує такої нормалізації як інші соціолекти (наприклад, наукова термінологія), декілька варіянтів однієї лексеми можуть співіснувати в живому мовленні без виникнення “конкуренції” і “забування” тих чи інших одиниць.

Значний вплив на варіятивність чинять не тільки внутрішньо-, а й зовнішньосистемні фактори. За Фішманом, існує цілий ряд соціолінгвістичних чинників, що впливають на мовний вибір.

1. Участь у групі (group membership) (Fishman, 1965). Перебуваючи у різних соціяльних оточеннях, мовці звертаються до різних лексичних варіянтів. Скажімо, у закладах вищої освіти спілкування ведеться літературним стандартом, вдома послуговуються побутовим стилем, що передбачає інакші мовні засоби. Утім під час занять на перервах студенти нерідко можуть відступати від мовного стандарту і використовувати сленг у міжосібних діялогах. Таким чином, самої лише участі студентів в академічній спільноті замало, аби детально схарактеризувати обставини їхнього мовного вибору.

2. Ситуація (Fishman, 1965). Ситуація – семантично доволі широкий спектр умов. До неї входять учасники, фізичні обставини, теми і функції дискурсу. Усі ці поняття для спрощення вивчення Фішман об’єднує одним поняттям *стилю*. Ситуаційні стилі, на думку Йооса, Лабова, Гумперца та інших (Fishman, 1965) належать до таких категорій:

- близькість – дистанція;
- солідарність – несолідарність;
- статусна рівність – нерівність;

Так, деякі стилі всередині кожної мови мовці індивідуально виокремлюють як індикатори більшої близости, неформальности, рівности тощо. Якщо ж йдеться про багатомовних людей, то вони не тільки вважають одну зі своїх мов більш регіональною, більш нестандартною, розмовною, а й асоціюють одну з мов з неформальністю, рівністю більше, ніж іншу.

Численні дослідження продемонстрували, що на мовний вибір впливають також стать, вік та соціяльно-політичні переконання.

Окрім лексичної варіятивности, має місце ще й семантична: те саме багатозначне слово може вживатися різними мовцями виключно в одному конкретному значенні. Так, Робінсон досліджувала семантичне варіювання лексеми *awesome, вивчаючи* мовлення 72 респондентів-англійців. Опитування продемонструвало, що різні вікові, соціяльні та гендерні групи вживають слово *awesome* у різному значенні. Вікова диференціяція використання однієї лексеми з різним змістом засвідчує історичну еволюцію семантики цього слова. Значення, у якому лексема вживається старшими поколіннями (*terrible*), є архаїчним і неактуальним; значення, до якого вдаються молоді (*great, impressive*), – новітнє і в майбутньому може витіснити інші зі словника (Robinson, 2010).

1.5. Словник обсценної лексики, укладений Катериною Бобровник

В одній зі своїх впливових праць Хомський вводить поняття мовної творчости (language creativity) – здатности людини створювати нескінченну кількість мовних одиниць, використовуючи скіченну кількість засобів (Chomsky, 1957). На прикладі обсценізмів в ролі словотвірних засобів виступають різноманітні морфеми, а продуктом мовної творчости стають кілька лексичних варіянтів. На мою думку, мовну творчість найкраще можна прослідкувати в мовленні. Саме мовлення є безпосереднім джерелом усіх лінгвістичних новотворів, тож воно містить найбільше матеріялу для аналізу актуального стану українських обсценізмів. Відповідно, для такого дослідження в нагоді стає словник Катерини Бобровник (2009), автоматично укладений на основі корпусу обсягом в 1,47 млн текстів твітів з соцмережі X (раніше – Twitter). У відібраних текстах за допомогою спеціяльно написаної програми К. Бобровник вишукувала словоформи, що містять принаймні одну значущу частину слова з укладеного нею списку обсценних квазікоренів та квазіоснов (Бобровник, 2019, с. 48).

У результаті такого добору було отримано словник, що складається із 7266 словоформ, які належать до таких категорій обсценної лексики:

1) лайка (*йооообана, отпиздить*);

2) жаргонно-нецензурні слова (*офігітільний, сіськіпіські, фігзнакатда*);

3) образливі слова за національністю (*жиди, негри, хачі*);

4) конкатеновані поєднання (*циганжид, хуюпіздрос, мудаїбізм*),

5) слова на позначення представників ЛГБТ (*гомосєкі, лохпідар*). (Бобровник, 2019, с. 48).

Словник Катерини Бобровник – це словник обсценних словоформ, зафіксованих у писемному мовленні. Тому, за словами його авторки, велика кількість словоформ пояснюється варіянтністю написання одного і того ж слова (Бобровник, 2019, с. 49). У словнику відсутня лематизація, усі словоформи подані алфавітним списком (рис. 1).

Рис. 1.1. Скриншот перших слів словника Катерини Бобровник

На рис. 1 видно, що словник містить словоформи, додані програмою через “хибне спрацьовування” (Бобровник, 2019): ctrlблять, jebaited, mtvговно. На мою думку, при опрацюванні словника такі словоформи не варто брати до уваги, оскільки вони, хоч і містять обсценні корені, не мають обсценного значення.

Висновки до першого розділу

У ході опрацювання теоретичної літератури я з'ясувала, що обсценна лексика є недостатньо вивченим, проте науково значущим шаром словникового запасу мови, дослідження якого стає дедалі актуальнішим в умовах сучасности. Зарубіжне мовознавство демонструє дещо вищий рівень зацікавленості цією проблематикою: вчені досліджують обсценізми в міждисциплінарному ключі — з позицій фонології, психолінгвістики та нейробіології, виявляючи навіть міжмовні закономірності у звуковій організації лайливих слів. У вітчизняній лінгвістиці системне вивчення обсценної лексики пов'язане передусім з іменем Лесі Ставицької, чиї праці заклали підґрунтя для подальших досліджень у цій галузі.

Аналіз класифікаційних підходів до українських обсценізмів засвідчив, що вони охоплюють як лексику сексуальної, так і скатологічної сфери. Кількісні підрахунки, здійснені Ставицькою на матеріялі сталих словосполучень, свідчать про домінування скатологічного типу в українській обсценній традиції, що дає підстави віднести її радше до *Shit*-культури, ніж до *Sex*-культури. Ця особливість, за припущенням науковиці, відображає специфіку ціннісної системи й табуйованих концептів українського мовного середовища.

Важливою теоретичною засадою дослідження є поняття мовної варіятивности, яка виявляється на фонетичному, морфологічному та лексичному рівнях. Для обсценізмів варіятивність є особливо виразною: нецензурних корені є доволі словотвірно продуктивними, унаслідок чого в мовленні постійно виникають нові форми та похідні. Писемне мовлення наочно фіксує ці процеси, до того ж досліджувати текст простіше, ніж аналізувати аудіозаписи мовлення, тому словник Катерини Бобровник, укладений на основі масиву текстів із соціяльної мережі X, є надзвичайно цінним джерелом для вивчення актуального стану українських обсценізмів та їхньої варіятивности. Саме цей матеріял становить одну з ключових емпіричних баз подальшого дослідження.

Також було проведено дослідження з аналізу корпусу дописів політичних діячів. Воно продемонструвало, що укладений словник може використовуватися

для автоматизованого аналізу текстових корпусів і виявлення обсценної лексики в сучасному публічному дискурсі. Водночас отримані результати вказують на необхідність подальшого розширення словникової бази за рахунок нових лексем, менш стилістично маркованих одиниць та графічних варіянтів обсценізмів.

РОЗДІЛ 2.
ЛЕМАТИЗАЦІЯ МАТЕРІЯЛІВ СЛОВНИКА КАТЕРИНИ БОБРОВНИК ЯК ЕТАП ЛЕКСИКОГРАФІЧНОГО ОПРАЦЮВАННЯ ОБСЦЕННОЇ ЛЕКСИКИ

2.1. Поняття лематизації в комп'ютерній лінгвістиці та засоби її реалізації

Деніел Джуравський надає таке визначення лематизації: "Лематизація – одна з частин нормалізації тексту, тобто його приведення [...] до стандартизованої форми. Це – визначення того, що декілька слів слова мають однаковий корінь, незважаючи на їхні поверхневі відмінності". Наприклад, слова *співає, співала* та *співатиме* є формами дієслова *співати*. Слово *співати* є спільною лемою цих слів, і лематизатор (програма, яка проводить лематизацію) перетворює всі ці слова на *співати*. Лематизація є важливою для обробки морфологічно складних, флективних мов, таких як українська, іспанська тощо (Jurafsky, Daniel & Martin, James, 2008).

Поряд із лематизацією виділяють також стемінг. Стемінг та лематизація – це методи попередньої обробки тексту в обробці природної мови (NLP). Зокрема, вони зводять кілька форм слів у текстовому наборі даних до одного спільного кореневого слова або леми.

Обидва методи особливо корисні в системах пошуку інформації, таких як пошукові системи, де користувачі можуть надсилати запит з одним словом (наприклад, *медитувати*), але очікувати результатів, які використовують будь-яку словоформу цього слова або спільнокореневі лексеми (наприклад, *медитує, медитація* тощо). Стемінг та лематизація також спрямовані на покращення обробки тексту в алгоритмах машинного навчання.

Дослідники дискутують щодо того, чи здатен штучний інтелект мислити в дійсності, і ця дискусія поширилася на комп'ютерну лінгвістику. Не має значення, чи можуть чат-боти та моделі глибокого навчання обробляти лише лінгвістичні форми, чи вони можуть розуміти семантику, незмінним є те, що

моделі машинного навчання повинні бути навчені розпізнавати різні слова як морфологічні варіянти одного базового слова. Навіть обробляти слова відповідно до морфології, а не семантики, для чого в нагоді стають стемінг та лематизація.

За Джуравським, стемінг – це спрощений процес лематизації, за якого відсікаються суфікси в кінці слова. Цей метод частково нагадує метод квазіфлексій, оскільки часто оперує не морфемами, а частинами слів, що не мають лінгвістичного значення. Так, головна відмінність стемінгу від лематизації полягає в тому, що базові форми лексеми (основа, аналог лем для лематизації) можуть не бути дійсними словами (Jurafsky, Daniel & Martin, James, 2008).

Для проведення стемінгу зручно використовувати бібліотеку NLTK. Одним з алгоритмів, що пропонує бібліотека, є стемер Портера. Це один з найпопулярніших методів стемінгу, запропонований у 1980 році. В його основі лежить ідея про те, що суфікси в англійській мові складаються з комбінації менших і простіших суфіксів. Цей стемер відомий своєю швидкістю та простотою. До основних застосувань стемера Портера входить інтелектуальний аналіз даних та пошук інформації. Однак його використання обмежене лише англійськими словами. Крім того, вихідна основа не обов'язково є значущим словом. Сам алгоритм досить довгий за своєю природою та є одним з найстаріших стемерів.

Більшість бібліотек, присвячених опрацюванню природної мови, базуються на словниках, зокрема й морфологічних. Написана за допомогою такої бібліотеки програма має впоратися з обробкою найуживаніших слів. Труднощі виникають тоді, коли вхідна лексема не є широко вживаною. Може виникнути припущення, що такими «складними» словами є лише неологізми, утім в дійсності лематизатор подеколи не в змозі опрацювати навіть відносно поширені власні назви, а в рідкісних випадках неправильно обробляє й загальні лексеми.

У ході навчання мені довелося працювати із бібліотекою Stanza. Stanza — це набір інструментів з відкритим кодом для обробки природної мови на Python,

що підтримує 66 мов. Порівняно з наявними широко використовуваними наборами інструментів, Stanza пропонує повністю мовно-незалежний нейронний конвеєр для аналізу тексту, включно з токенізацією, розширенням багатослівних токенів, лематизацією, позначенням частин мови та морфологічних ознак, розбором залежностей та розпізнаванням іменованих сутностей.

Іншою бібліотекою, що стає в нагоді при лематизації, є pymorphy2. За словами розробників, вона здатна:

- лематизувати словоформи;
- змінити граматичну форму слова (наприклад, змінити рід прикметника);
- надати інформацію про граматичні характеристики вхідного слова (частина мови, рід, число, відмінок та інші).

В основі бібліотеки лежить словник OpenCorpora – словник російської мови, у якому представлено 391845 лем. Власне, у цьому полягає головний недолік pymorphy2. Перша версія бібліотеки – pymorphy – була розрахована для обробки виключно російської мови. pymorphy2 уже містить пакет для української, але він, як зазначають самі розробники на своєму сайті, є експериментальним.

2.2. Проблеми лематизації рідковживаних лексем.

У ході навчання за допомогою бібліотеки Stanza я опрацьовувала дві вибірки: наукового (підручник “Вступ до мовознавства” Михайла Кочергана) та художнього (роман “Танго смерті” Юрія Винничука) стилів. В обидвох вибірках програма припускалась помилок не тільки у встановленні леми за словоформою, а й у приписуванні частини мови.

Таблиця 2.1.

Приклади хибного спрацювання автоматичної лематизації

Словоформа	**Лема за Stanza**	**Правильна лема**
ф	ф	Фердинанд
однина	одниний	однина
математиці	математиця	математика
мовці	мовка	мовець

лексем	лекс	лексема
луна	луно	луна
словоформ	словоф	словоформа
балюстради, балюстраду, балюстраду	балюстрад	балюстрада

У таблиці 1 наведено декілька прикладів хибного спрацювання лематизатора Stanza. Часто викликали проблеми через скорочення, оскільки програма не враховує контекст і не може відтворити повне слово за однією літерою. Іноді помилки траплялися з науковою або рідковживаною лексикою (*лексема, словоформа, балюстрада*). Скоріш за все, цих слів не було в словнику, на якому тренувалась Stanza, тому програма намагалась лематизувати лексику за аналогією до інших слів, що мали такі ж гіпотетичні закінчення. Програма не впоралась із лематизацією таких слів як *однина, математиці, мовці* та *луна*. На мою думку, це пов'язано з тим, що Stanza або не завжди правильно встановлює частину мови, або, хоч і визначає частину мови правильно, але має певні проблеми із визначенням роду. Відповідно, програма приписує словоформам неправильні парадигми та, згодом, неправильні леми.

Попередній досвід роботи зі Stanza допоміг краще спланувати етапи лематизації матеріялу зі словника К. Бобровник, які я описуватиму далі. Попри хибні спрацювання бібліотеки, вона все ж може надати прийнятний результат і, на відміну від російської бібліотеки pymorphy2, окремо підтримує українську мову. Так, маючи досвід роботи зі Stanza, та врахувавши недоліки pymorphy2, серед двох бібліотек для роботи я обрала першу.

2.3. Автоматична лематизація словника та визначення частин мови.

Далеко не всі одиниці, наведені у словнику Катерини Бобровник, є лексемами, отже при обробці словника їх варто було виключити. Застосунків чи бібліотек, які б автоматично видалили хибні входження словника, не існує, тому цей етап роботи довелося виконати вручну. Я власноруч вичитала увесь

матеріал, видаляючи несловесні одиниці та редагуючи деякі слова чи словосполучення. Так, наприклад, зі словника були вилучені такі входження як *всьопіздєцвам, нахуйнікомунепотрібних*. У свою чергу, слова на кшталт *йобтвоюмать, йопсєльмоксєль* були розділені пробілами чи дефісами й залишені в словнику у відредагованому вигляді як *йоб твою мать, йопсєль-моксєль*.

Загалом словоформи, що підлягали видаленню чи редагуванню, можна поділити на такі категорії:

1) слова з повторенням графем (*ааааафігєть, бляхаааааааааааааааа, курвоо*);
2) багатослівні одиниці – лексеми, які складалися з двох чи більше слів і не були сталими сполуками (*акустикусраку, всьопропаловсьоплоховсємупіздєц, їбалисядикігуси* – видалила; *їбать-копать, йопсєль-моксєль* – лишила);
3) обсценні оказіоналізми – слова з обсценними коренями, які не є поширеною лайкою, а були створені мовцем ситуативно (*ахренологи, банкихуянки, бляблятека, журнабляді*);
4) необсценні одиниці – слова, в яких обсценний корінь збігся з необсценним набором графем (*бакалявр, вироблять*);
5) деякі лексеми з правописом, що непритаманний українській мові або є просторічним (майже всі слова з коренем *єб* видалила або замінила на *їб*, *афігєнний → офігенний*, *їбаця → їбатись*).

Останній пункт був вимушеним кроком. Оскільки джерельний словник створений на базі природного письмового мовлення, варіятивність деяких лексем може бути надзвичайно великою. Так, наприклад, у словнику К. Бобровник зібрано 500 словоформ із коренем *єб* та 662 словоформи із коренем *їб.* Слова, що містять два різні корені, не можуть повністю збігатися за іншими морфемними характеристиками через різну кількість одиниць у двох вибірках, а також через природу самого словника, утім більшість словоформ все ж структурно ідентичні. У майбутньому при створенні словника на основі такого списку лексем значна кількість варіянтів приведе до утворення багатьох статей,

де лексеми матимуть однакове тлумачення та частиномовну належність, але відрізнятимуться однією чи двома літерами. Тому було прийнято рішення залишити лише словоформи із графічним записом, притаманним українській фонетиці (із коренем *їб*, наприклад), та вилучаючи ті, де траплялись записи, більш властиві російській фонетиці (тут – із коренем *єб*). Якщо у списку словоформ зі зросійщеним записом були такі, що не мали власних відповідників у списку словоформ із "нормативним" записом, то зросійщена морфема замінювалась на притаманну українській морфеміці, і словоформа залишалася у словнику. У підсумку кількість слів у словнику скоротилась від 7266 до 4081.

Після ручної обробки була написана програма для автоматичної лематизації словникових одиниць. Для вирішення цієї задачі було вирішено використати бібліотеку Stanza. Як згадувалось раніше, хоча Stanza серед багатьох інших мов підтримує зокрема українську, вона іноді припускається помилок у лематизації загальновживаної лексики. Враховуючи цю обставину, можна було очікувати ще більшої кількости похибок під час опрацювання такого специфічного типу лексики як обсценізми. Бібліотека, насправді, відносно вдало лематизувала більшість одиниць, та водночас припустилась помилок. Деякі з них наведені у таблиці 2. Знову простежується така тенденція: Stanza неправильно визначає частину мови, а отже й рід і парадигму словоформи, у результаті приписує некоректну лему.

Таблиця 2.2

Приклади хибного спрацювання лематизації обсценізмів
за допомогою Stanza

Словоформа	**Лема за Stanza**	**Правильна лема**	**Остаточна лема**
ахєрєлі	ахєрєль	ахєрєлі	охеріли
всрана	всран	всраний	всраний
грьобаній	грьобанії	грьобаний	грьобаний
засри	заср	засрати	засрати
зассалі	зассаль	зассалі	зассали
напиздів	напизед	напиздіти	напиздіти

Після ручного редагування отриманого списку лем кількість словникових одиниць знов зменшилась майже вдвічі: з 4081 слова залишилось 2250.

Далі за допомогою тієї ж бібліотеки Stanza було написано програму, яка визначила частиномовну належність кожної леми. Як уже згадувалось, бібліотеці складно виконувати цю операцію, тому й на цьому етапі виникла необхідність ручного редагування. У таблиці 3 наведені деякі випадки хибного спрацьовування програми.

Паралельно із визначенням частин мови програма вилучила зі списку лем усі повтори, також було проведено повторне редагування словника. У результаті було отримано остаточний реєстр обсягом 1146 одиниць. Кількісне співвідношення частин мови продемонстроване у таблиці 4.

Таблиця 2.3

Приклади помилок автоматичного визначення частини мови лексеми за допомогою Stanza

Лема	**Частина мови за Stanza**	**Правильна частина мови**
недотрах	вигук	іменник
хуйло	прислівник	іменник
гімно	прислівник	іменник
охуєнськи	прикметник	прислівник
херзна-шо	прикметник	займенник
похер	дієслово	прислівник
полужопіє	дієслово	іменник
хуїти	іменник	дієслово
хера собі	іменник	вигук

Таблиця 2.4

Кількісне співвідношення частин мови у словнику Катерини Бобровник

Частина мови	Кількість	Приклад
Іменник	430	*фігня*
Дієслово	368	*офігівати*
Прикметник	171	*офігезний*
Прислівник	81	*пофіг*
Вигук	51	*фігак!*
Займенник	25	*фігзна-який*
Числівник	11	*дофіга*
Сталий вираз	9	*фіг тобі*
Усього	**1146**	

Найбільшу кількість лем становили іменники, дієслова, прикметники та прислівники, що загалом відповідає загальній тенденції частотності частин мови серед загальновживаної лексики. Характерною особливістю обсценізмів виявилось кількісного переважання вигуків над займенниками та числівниками. Тимчасом як у нормативній лексиці вигуки посідають нижчі місця у частотних списках, серед лайки вони значно більш поширені. Імовірно, це пов’язано з тим, що джерельною базою словника Катерини Бобровник слугувало живе мовлення розмовного стилю, у якому кількість вигуків вища, ніж, скажімо, в публіцистичних текстах. До того ж, на мою думку, переважання вигуків над деякими типово частотнішими частинами мови можна вважати відмінною рисою обсценної лексики.

Висновки до другого розділу

У другому розділі було розглянуто теоретичні та практичні аспекти лематизації як ключового методу нормалізації тексту в обробці природної мови. З'ясовано, що лематизація, на відміну від стемінгу, зводить словоформи не до

умовного кореня, а до словникової форми — леми, що робить її точнішим інструментом для лексикографічної роботи. Обидва методи є важливими для систем пошуку інформації, однак лематизація є більш прийнятною для завдань, що передбачають подальший словниковий опис матеріялу.

У ході порівняльного аналізу доступних інструментів для автоматичної обробки українськомовних текстів було встановлено, що бібліотека Stanza є найбільш придатною для цілей дослідження, оскільки, на відміну від pymorphy2, забезпечує окрему підтримку української мови. Водночас виявлено суттєві проблеми автоматичної лематизації: бібліотека припускається помилок при обробці скорочень, рідковживаної та наукової лексики, а також некоректно визначає частиномовну належність окремих одиниць, що у свою чергу призводить до хибного визначення леми.

Практична робота зі словником Катерини Бобровник засвідчила, що автоматична обробка обсценної лексики потребує обов'язкового ручного редагування на кількох етапах. На першому етапі з початкового списку у 7266 словоформ було вилучено несловесні одиниці та графічні варіянти з ознаками російської фонетики, що скоротило матеріял до 4081 одиниці. Після автоматичної лематизації та подальшого ручного редагування отриманих результатів кількість словникових одиниць зменшилась іще майже вдвічі — до 2250 лем.

Визначення частиномовної належности лематизованих одиниць також вимагало ручного втручання через систематичні помилки бібліотеки. Аналіз частиномовного складу отриманого словника загалом відповідає загальномовним тенденціям: найчисельнішими групами виявились іменники, дієслова, прикметники та прислівники. Характерною особливістю обсценної лексики стало помітне переважання вигуків над займенниками та числівниками, що, ймовірно, пов’язано із експресивною функцію нецензурної лексики в мовленні.

РОЗДІЛ 3.
УКЛАДАННЯ БАЗИ ДАНИХ ТА СТВОРЕННЯ ЗАСТОСУНКУ ДЛЯ СЛОВНИКА

3.1. Добір лексикографічних матеріялів та підготовка до створення бази даних

Після опрацювання словника К. Бобровник було прийнято рішення погрупувати отримані лексеми довкола певного інваріянта-кореня. Існує значна кількість слів, які мають ідентичну семантику та водночас подібне графічне вираження (*відпиздюлений, однижджений, однижжений, відпизжений, відпижджений, відпижжений, пижджений* або *охріненний, охрінезний, охрінітельний, охрінітительний*). Розведення таких лексем як окремих самостійних реєстрових слів призвело б до повторів ідентичних за тлумаченням статей. Обсценній лексиці також властива велика кількість синонімічних рядів. Наприклад, у кінці свого словника, де Леся Ставицька подає покажчик синонімів, 22 лексеми означають “коханка”, 44 слова мають значення “статевий акт”, 196 слів означають “жіночий статевий орган” (Ставицька, 2008, с. 426). Часто такі лексеми не мають виразних стилістичних відтінків, що виділяли б їх з-поміж інших. Їм властивий відносний ступінь згрубілости або пестливости, за яким пересічний читач словника міг би виділити “більш” чи “менш” відверті форми, але загалом такі слова мають ідентичні тлумачення. Відповідно, враховуючи значну синонімію та велику кількість аналогічних лексем із незначними відмінностями у морфологічній будові або фонетичному складі, я прийняла рішення об’єднати такі слова під однією дефініцією. Для того, щоб словник усе ж мав вигляд тлумачного словника, у таких групах слів було обрано лексему-корінь, яка відігравала роль реєстрового слова словника. Принципи вибору реєстрового слова (інваріянта-кореня, лексеми-кореня) докладніше описані нижче.

Так, наприклад, до однієї групи могли увійти різні графічні варіянти одного слова (*бл, блять, блт, бля; блядський, бляцький*).

Іменники були об’єднані у спільні групи разом із відповідними зменшено-пестливими або згрубілими формами до них (*блядина, блядинка, блядище*). У групах прикметників поєднувались спільнокореневі слова із різним ступенем вияву ознаки (*хріновий, хріноватий*). Групи дієслів могли містити такі лінгвістичні параметри:

1) дві форми інфінітиву (*дристати, дристать*);
2) дві форми інфінітиву зворотних дієслів (*наїбнутися, наїбнутись*);
3) одне й те саме дієслово у двох видах (*задобвати, задовбувати*);
4) одне й те саме дієслово у різних фазових значеннях дії (*задовбати, задовбувати, позадовбувати*):
5) одне й те саме дієслово у різних кількісних аспектах виконання дії:

а) за кількістю самої дії – однократної чи багатократної (*спиздонути, спиздіти*);

б) за інтенсивністю дії (*вийобуватися, повийобуватися*);

Фразеологічні звороти та сталі вислови також були об’єднані за інваріянтом. У такі групи входили вислови, що відрізнялися певним словом (*сракою чути, жопою чути*), варіянтом певного слова (*йоп тудей, йопс тудей, йопстудей; їбати мозок, єбати мозок, їбать мозок, єбать мозок*) чи правописом (*камінаут, камін-аут, камінг аут*). Варіятивність фразеологізмів та сталих висловів була здебільшого спричинена варіятивністю частин мови, принцип групування яких описаний вище.

Після обговорення попередньої роботи зі словником було переглянуто рішення з редагування словника та урізання “невластивих” українській мові лексем. Оскільки головною мовою “Словника обсценізмів та сленгу” є повне представлення живого сучасного мовлення, для кожного слова, незалежно від частиномовної належности, вручну було додано можливі фонетичні варіянти (*пизда, пізда, писда, пісда, песда, пєзда, пєсда*). Рішення розширити словник було прийнято також з урахуванням того, що фонетична (і, як наслідок, графічна) варіятивність є однією з головних характеристик обсценізмів. Окрім цього, для того, щоб “Словник обсценізмів та сленгу” приніс користь не тільки звичайним

користувачам, а й лінгвістам-інженерам, розширення реєстрового списку словника є обов'язковою умовою. Такий список обсценізмів можна використати для організації автоматичного пошуку відповідних лексем у будь-якому тексті, автоматичній заміні нецензурної лексики, визначення рівня токсичности тексту абощо. Відповідно, обсяг такого словника прямо пропорційно впливає на якість лінгвістичного дослідження. Що більше лексем врахує реєстровий список словника, то вищою буде точність програми з обробки природної мови. Таким чином, врахувавши фонетичні варіянти у "Словнику обсценізмів та сленгу", я зберегла важливу рису досліджуваного пласту лексики та забезпечила гнучкість у використанні словника для різної цільової авдиторії.

Обсценна лексика є ненормативною, з чого випливає її ненормованість. У багатьох обсценізмів немає фіксованих словником початкових форм, а кількість фонетичних та графічних варіянтів доволі висока. Тому у ході опрацювання матеріялів словника Катерини Бобровник я власноруч обирала форму на позначення інваріянта (реєстрове слово) за такими принципами:

- для іменників іваріянтами були лексеми без зменшено-пестливих суфіксів або суфіксів згрубілости;

- для прикметників варіянт лексеми з найменш вагомим стилістичним забарвленням/інтенсивністю вияву ознаки;

- для дієслів варіянт лексеми недоконаного виду;

- якщо слово є запозиченням з російської і написане згідно з російською фонетичною системою, то воно залишиться варіянтом, натомість інваріянтом стане та ж лексема, написана відповідно до української вимови;

- частота вживання слова у дописах соцмережі X (Twitter).

Останній критерій був одним із найскладніших для перевірки. Я вирішила досліджувати частоту вживання слова у соцмережі X (Twitter), оскільки на основі дописів звідти Катерина Бобровник створила свій словник. Цей метод використовувався не часто, оскільки проведення подібного дослідження за допомогою програмних засобів є тривалим у часі й наразі потребує додаткових зусиль. Останні кілька років доступ до постів у соцмережі X (колишній Twitter)

був обмежений внутрішніми правилами платформи, оскільки зовнішні організації намагалися скрепйити дані соцмережі з надмірною інтенсивністю. (Bitdefender, 2023) Обмеження на завантаження інформації з X стали на заваді звичайним розробникам також, тому, за відсутності можливості проведення якісного аналізу програмними засобами, я проводила перевірку вручну. Хоча цей метод потребував значної кількости часу, у випадках із певними лексемами він давав доволі показові результати. Так, наприклад, слово *зассаний* у період 2024-2026 рік вжили 22 рази, тим часом як *засцяний* лише за 2025-2026 роки вжили понад 70 разів. Крім критерію частотности, у цьому випадку також спрацьовує інший фактор. На відміну від лексеми *засцяний*, слово *зассаний* є властивим російській мові і має відповідні фонетичні характеристики, відображені на письмі. Відповідно, головним словом у словниковій статті було обрано лексему *засцяний*.

Після опрацювання словника Катерини Бобровник я перейшла до роботи зі словником Лесі Ставицької.

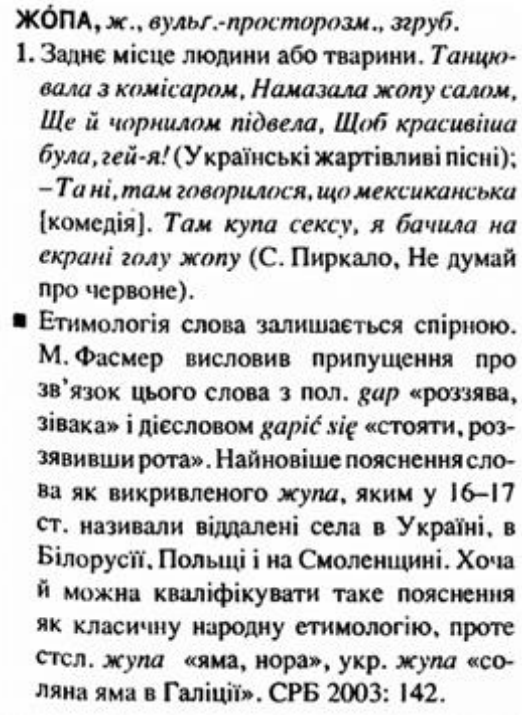
Словник Лесі Ставицької "містить точний опис значень близько 5000 слів і стійких словосполучень, подає їх стилістичні характеристики, довідки про походження та історико-культурний коментар, а також фіксує випадки нестандартного, «ігрового» вживання"(Ставицька, 2008). Словникова стаття словника має таку структуру:

- реєстрове слово;
- граматична інформація (частина мови, рід для іменників та вид для дієслів);
- експресивно-стилістична характеристика;
- тлумачення;
- ілюстративний матеріял.

Якщо реєстрове слово має фонетичні чи регіональні варіянти, вони подані в дужках після нього. Ця характеристика була корисною для створення мого словника, оскільки допомогла розширити списки варіянтів реєстрових одиниць, відібраних зі словника Катерини Бобровник. Окрім варіянтів, у частині з

граматичною інформацією подаються словозмінні закінчення прикметників та позначаються невідмінювані частини мови.

Після граматичної характеристики слова йде низка позначок, призначення яких – максимально повно охарактеризувати функціональне (*арг., жарг., фольк., розм., просторозм., обсц.* та ін.), емотивно-експресивне вживання (*згруб., вульг., лайл., жарт.-ірон. та ін.*), а також закріпленість за діялектами та говорами української мови (Ставицька, 2008, с. 75). У дужках після *жарг.* даються додаткові позначки, які уточнюють жаргонний соціолект: *жарг. (мол.) – у молодіжному жаргоні* або вказують на поширення слова у жаргонізованій розмовній мові (у загальному сленгу): *жарг. (жрм.)*.(Ставицька, 2008, с. 75)



ЖО́ПА, *ж., вульг.-просторозм., згруб.*

1. Заднє місце людини або тварини. *Танцювала з комісаром, Намазала жопу салом, Ще й чорнилом підвела, Щоб красивіша була, гей-я!* (Українські жартівливі пісні); *– Та ні, там говорилося, що мексиканська* [комедія]. *Там купа сексу, я бачила на екрані голу жопу* (С. Пиркало, Не думай про червоне).

■ Етимологія слова залишається спірною. М. Фасмер висловив припущення про зв'язок цього слова з пол. *gap* «роззява, зівака» і дієсловом *gapić się* «стояти, роззявивши рота». Найновіше пояснення слова як викривленого *жупа*, яким у 16–17 ст. називали віддалені села в Україні, в Білорусії, Польщі і на Смоленщині. Хоча й можна кваліфікувати таке пояснення як класичну народну етимологію, проте стсл. *жупа* «яма, нора», укр. *жупа* «соляна яма в Галіції». СРБ 2003: 142.

Рис. 3.1. Приклад словникової статті зі словника Лесі Ставицької

Тлумачення в словнику добиралися з різних джерел. Деякі дефініції в готовому вигляді бралися з українських та закордонних джерел, деякі розроблялися з урахуванням семантичних параметрів лексикографічної бази, яка міститься у різнотипних словниках, інші розроблялись як власні тлумачення із залученням на явних контекстів (Ставицька, 2008, с. 76). У деяких випадках правильність дефініції перевірялась через опитування інформантів.(Ставицька, 2008)

Окрім вже описаних даних, словникова стаття містила також історико-етимологічний коментар, який у деяких випадках включав розлогий опис походження реєстрового слова (див. слово Жопа я потім пронумерую картинки). Наприкінці словникової статті також часто подаються синоніми до даного реєстрового слова. З усієї лінгвістичної інформації, яку надає словник Лесі Ставицької, я сфокусувалася на двох частинах: реєстрове слово та тлумачення. У ході роботи з лексикографічною працею я відбирала обсценні лексеми, яких не трапилося в словнику Бобровник і які, відповідно, не були зафіксовані в реєстрі "Словника обсценізмів та сленгу". Також добирала тлумачення для слів, які вже містилися у реєстрі словника К. Бобровник. Якщо до лексеми, попередньо відібраної зі словника К. Бобровник, дописувалось тлумачення зі словника Л. Ставицької, то серед її джерел у "Словнику обсценізмів та сленгу" вказувалися дві праці: словник Катерини Бобровник та словник Лесі Ставицької. Таким чином, реєстр "Словника обсценізмів та сленгу" було поповнено новими словами та визначеннями. Відповідно, інформацію про приклади-ілюстрації, функціональну регіональну, та стилістичну характеристики лексем, наведені у статтях Лесі Ставицької, не було враховано у власному словнику.

Після того, як "обсценна" частина "Словника обсценізмів та сленгу" була заповнена новою лексикою або її варіантами та поповнилась дефініціями Лесі Ставицької, у робочій таблиці комірки для тлумачення більшості слів залишились пустими. З метою удосконалення фінального словника такі дефініції були прописані вручну, з урахуванням лексикографічного стилю та семантичних параметрів лексикографічної бази, яка містилася у словнику Ставицької.

Деякі знайдені лексеми виявились неосемантизмами – словами зі звичною формою, але новим значенням (Цигвінцева Ю. О, 2025). Явище неосемантизації не є рисою, властивою виключно обсценному прошарку лексики. Зміна семантики слів є природним лінгвістичним процесом, про що свідчить велика кількість праць на цю тему. В українському мовознавстві неосемантизми вивчали Ю. Цигвінцева (Цигвінцева, 2025), (Цигвінцева, 2020), (Цигвінцева, 2024), О. Білецька (Білецька, 2025), Л. Кислюк (Кислюк, 2025), (Клименко,

Карпіловська, Кислюк, 2008), Н. Клименко, Є. Карпіловська (Клименко, Карпіловська, Кислюк, 2008). За кордоном це явище досліджували Robinson (Robinson, 2010), Geeraerts (Geeraerts Dirk, Speelman Dirk, Heylen Kris, Montes Mariana, De Pascale, Stefano Franco, Karlien Lang, 2024), Keidar (Daphna Keidar, Andreas Opedal, Zhijing Jin, and Mrinmaya Sachan, 2022).

Останні кілька років можна спостерігати значний вплив війни на функціонування української мови (Кислюк, 2025). Через зміну реалій підвищується частота вживання власне воєнної лексики. Також у мові актуалізуються слова та словосполучення для номінації предметів чи явищ нової дійсности (*блокпост, обстріл, повітряна тривога, окупація, воєнний злочин*) (Кислюк, 2025). Значний пласт неосемантизмів пов'язаний із воєнним станом та вже згаданою потребою створення лексикону, який би влучно описував навколишню реальність.

Рис. 3.2. Неосемантизація слова блядіна

У “Словнику обсценізмів та сленгу” однією з неосемантизованих лексем є слово *блядіна* у значенні *російська ракета, яка летить на територію України*. Уперше цю лексему в новому значенні було вжито в телеграм-каналі «Хуйовий Харків» 6 березня 2022 року (Шевченко, 2022). Новітнє тлумачення було перенесено з новотвору-пароніма *блєдіна*.

Деякі користувачі стверджують, що лексеми *блєдіна* та *блядіна* (*блядина*) мають різні значення (Шевченко, 2022), проте у ході аналізу постів соцмережі X я виявила доволі частотне вживання слова *блядіна* у значенні ворожої ракети (рис. 3.2).

Рис. 3.3. Графічний мем, узятий із соцмережі X (коментар користувача @boorn7), який демонструє значення новотвору блєдіна

Довкола походження лексеми *блєдіна* точаться суперечки. Згідно з даними Urban Dictionary, ця лексема вперше виникла у колі львівських волонтерів та набула популярности у соцмережі X (колишній Twitter) (Urban Dictionary, 2024). Слово почали використовувати у ході роботи із гуманітарною допомогою з-за кордону. Волонтери помітили фонетичну подібність між неологізмом та назвою бренду Bledina – провідного французького бренду дитячого харчування, що належить компанії Danone. Так, за словами користувачів X (колишній Twitter),

лексема поширилася в колах волонтерів, а згодом набула популярности у соцмережах (Шевченко, 2022). Важливу роль у поширенні цього слова взяла львівська активістка Меланія Подоляк після того, як вжила його у своєму пості в X.

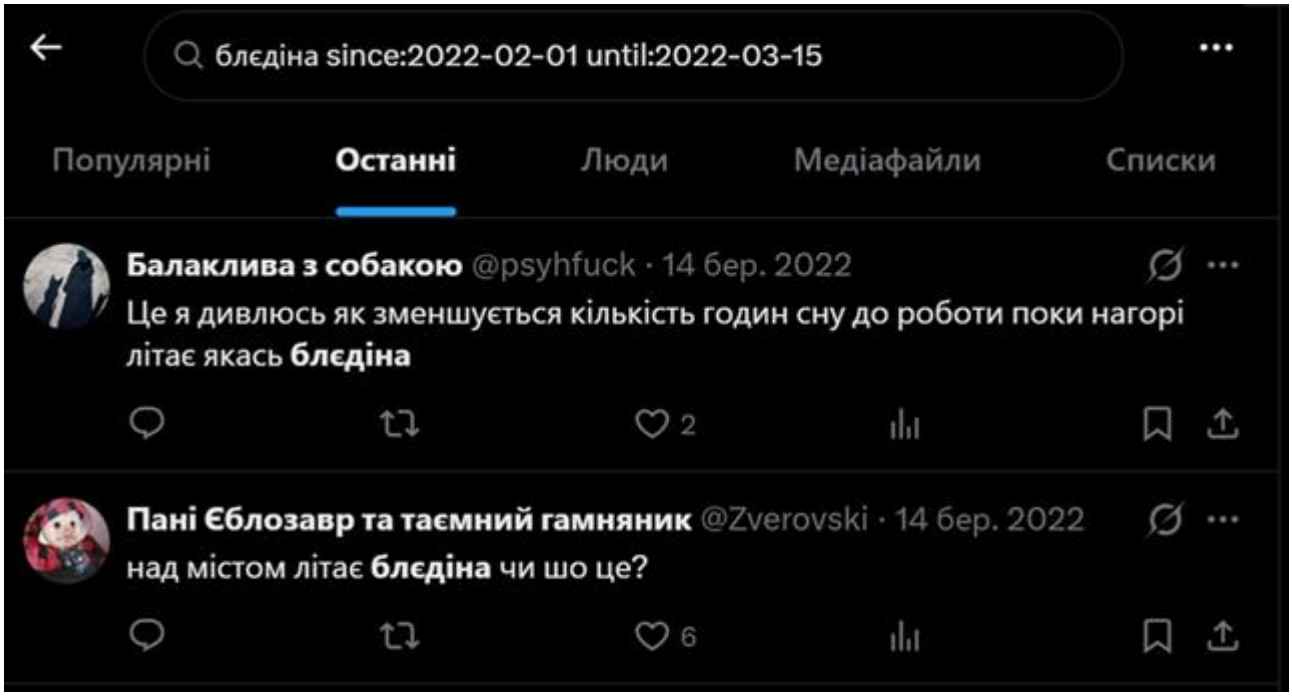


Рис. 3.4. Згадки слова блєдіна *у проміжку від 1 лютого 2022 до 15 березня 2022*

Нещодавно у соцмережі X виникла дискусія, яка ставить під сумнів попередню версію виникнення лексеми *блєдіна*. Пост, у якому Меланія Подоляк вперше згадує марку дитячого харчування під час повітряної тривоги (Podolyak, 2022), датується 29 березня 2022 року. Користувачі тимчасом зауважують, що в соцмережі існують записи із новотвором, які з'явилися дещо раніше. Наприклад, користувач @eighteen_th зазначає, що перші пости зі згадкою цього слова в X з'явилися 14 березня і були створені іншими користувачами, без посилання на бренд дитячого харчування (рис. 3.3) (eighteen_th, 2024). Це дослідження також можна поставити під сумнів, оскільки *блєдіна* має фонетичний варіянт – *бледіна*. Тому я самостійно провела пошук постів зі словом *бледіна* за той самий часовий проміжок, але не знайшла жодного допису зі згадкою цієї лексеми.

Інші джерела згадують пост за 13 березня 2022 зі згадкою слова блєдіна (Шевченко, 2022), утім наразі ані допис, ані акаунт користувачки-авторки, знайти в соцмережі неможливо.

Таким чином, походження лексеми блєдіна досі є дискусійним питанням. Для глибшого аналізу питання можна провести опитування спільноти львівських волонтерів, які, за здогадками, безпосередньо брали участь у створенні та поширенні новотвору. Перш ніж з'явитися в медіапросторі, слово ймовірно виникло в усному спілкуванні, тому таке дослідження дозволило б підтвердити або спростувати першу появу лексеми саме у колі львівських волонтерів. Усе ж за інформацією, доступною в Інтернеті, тобто за писемними згадками, зв'язок походження слова *блєдіна* із брендом дитячого харчування викликає певні сумніви, хоча не можна заперечити той факт, що Меланія Подоляк, як медійна особа, зробила значний внесок у поширення даної лексеми.

Рис. 3.5. Неосемантизація лексеми відпердолити

Іншим прикладом часткової неосемантизації було слово *віднердолити*. За статтею Л. Ставицької, воно має лише одне значення: *здійснити статевий акт*, проте у ході мого дослідження соцмережі Х я виявила велику кількість постів, де дана лексема була вжита у значенні *прибирати* (рис. 3.5).

Для таких лексем, чия семантика змінилася частково, або тих, які зазнали неосемантизації, тлумачення були змінені або створені власноруч відповідно.

Я вирішила не обмежуватися виключно добором обсценізмів, оскільки маю бажання створити словник із лексикою, що якомога краще відповідає сучасним реаліям. Тому для доповнення моєї лексикографічної праці я обрала саме пласт сленгізмів, оскільки сленговий лексикон – один із найбільш рухливих пластів мови, схильний до примх моди, зміни мовного смаку (Должникова, 2013).

Сленгізми було дібрано із вебсайту SlangZone. SlangZone — це словник живої мови, який поповнювати може кожен користувач (SlangZone, 2026). Метою словника є збір ненормативної лексики, а саме: “сленгу, діалектизмів, неологізмів, обсценізмів, вульгаризмів” для збереження мовного багатства (SlangZone, 2026). Електронний словник не містить визначень у класичному лексикографічному стилі, оскільки всі статті створюються користувачами без модерації. Щоб взяти участь у наповненні сайту, потрібно пройти коротку реєстрацію за електронною поштою Google або через власний акаунт в Х (колишній Twitter). Через відсутність кваліфікованого редактора та простоту процесу додавання лексеми, статті часто можуть повторюватися. Так, наприклад, на сайті містяться три статті за словом *вайб-чек* (*вайбчек* або *вайб чек*) з подібними тлумаченнями та дві статті про слово *вайб-чекати* (*вайб-чекнути* та *вайбчекнуть*). Сайт підтримує 50 мов, серед яких українська, англійська, німецька, естонська, фінська, угорська, грецька, албанська, узбецька, казахська, грузинська, іврит, індонезійська, тайська, в’єтнамська та японська. Станом на травень 2026 року в україномовному словнику SlangZone міститься 2524 статті із урахуванням повторів. Усі ці лексеми були додані 65 користувачами. Найбільша кількість статей, вписана одним учасником, становить 1555 слів, а

найменша – 1 слово. Хоча модерація у словнику відсутня, користувачам надається можливість оцінити якість статті. Наприкінці кожної статті є дві кнопки – зображення великого пальця вгору та зображення великого пальця вниз. Поряд із кнопками відображається кількість схвальних та негативних оцінок. За допомогою такого голосування користувач, що ознайомлюється зі статтею, може оцінити, наскільки її вміст відповідає думці решти користувацької спільноти.

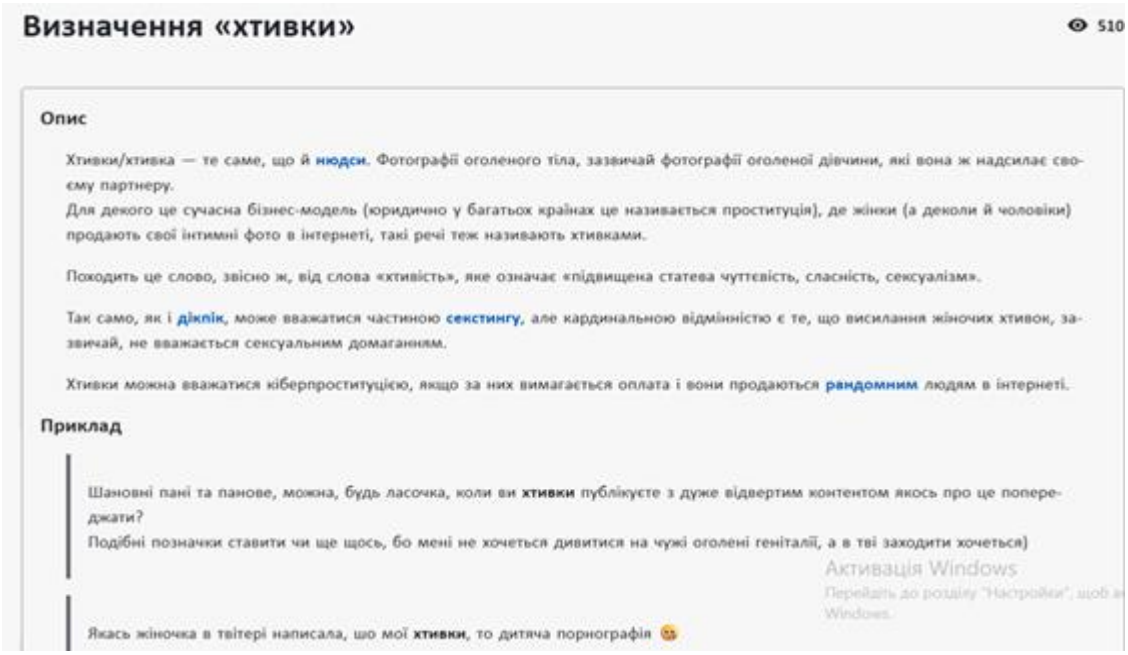


Рисунок 3.6. Приклад статті з вебсайту SlangZone (ч. 1)

Кожна стаття обов'язково містить реєстрову одиницю, тлумачення та приклад-ілюстрацію. Деякі автори також додають історію походження лексеми, візуальні матеріяли для унаочнення та гіперпосилання на пов'язані із даним словом статті.

Для свого словника із вебсайту SlangZone я добирала реєстрове слово та тлумачення. Дефініції, додані користувачами, не відповідали лексикографічному стилю, але окреслювали загальну семантику слова. Такі тлумачення не могли бути включені до мого словника, але стали основою для фінальних дефініцій. Усі тлумачення, відібрані зі словника SlangZone були скорочені та відредаговані з метою збереження простоти форми та чіткості змісту, і в такому вигляді увійшли

до словника. Більшість відібраної лексики складали сленгізми-запозичення з англійської, які виникли у соціяльних мережах або ігровому середовищі.

Рисунок 3.7. Приклад статті з вебсайту SlangZone (ч. 2)

3.2. Укладання бази даних “Словника обсценізмів та сленгу”

Після опрацювання трьох джерел я наповнила лексикографічною інформацією електронну таблицю в Google Sheets. Вона містила такі колонки:

1) Слово* – реєстрова одиниця статті «Словника обсценізмів та сленгу»; якщо лексема має більше, ніж один варіянт, вона повторюється у колонці стільки разів, скільки варіянтів (або, в деяких випадках, прикладів вживання) має це слово;
2) Варіянт* – варіянт реєстрової одиниці словникової статті; кожному слову приписувався щонайменше один варіянт – ця сама лексема;
3) Сфера вживання – вказується сфера, з якої походить дане слово (ця колонка була актуальною лише для сленгізмів, де в словникових статтях користувачі подавали етимологію реєстрових слів);
4) Відповідник – вказується український відповідник лексеми (ця колонка також заповнювалась лише для окремих сленгізмів-запозичень з англійської);

5) Тлумачення* – подається дефініція зі словника Л. Ставицької, власне написане тлумачення або відредаговане тлумачення зі словника SlangZone;
6) Приклад вживання – подається приклад вживання у письмовому мовленні (лише для деяких лексем; приклади відібрані із соцмережі X (колишній Twitter));
7) Джерело 1* – вказується словник, із якого було дібрано лексему або лексему та тлумачення;
8) Джерело 2 – вказується словник, із якого було дібрано тлумачення.

Колонка *Джерело 2* була створена для лексем, які були знайдені в словнику К. Бобровник та словнику Л. Ставицької. Зірочкою (*) в переліку колонок було позначено ті поля, які були обов'язково заповнені для всіх реєстрових одиниць.

A	B	C	D	E	F	G	H
Слово	**Варіянт**	**Сфера вживання**	**Відповідник**	**Тлумачення**	**Приклад вживання**	**Джерело 1**	**Джерело 2**
срати	срати			1. Випльонути кишечник; випорожнюватися (про людей і тварин). 2. Дуже боятися. 3. Зневажати кого-, що-небудь, з презирством ставитися до когось, чогось.		Бобровник	Ставицька
апіздати	апіздати			Запізнитися.		Бобровник	Ставицька
офіг	офіг			Стан сильного здивування, шок.		Бобровник	
офіг	афіг			Стан сильного здивування, шок.		Бобровник	
ахуй	охуй			Стан сильного здивування, шок.		Бобровник	
балда	балда			Чоловічий статевий орган великого розміру.		Ставицька	
бендюжний	бендюжний			Огидний, нечесний.		Ставицька	
бздо	бздо			Шлункові, кишкові гази.		Бобровник	Ставицька
бздо	бздух			Шлункові, кишкові гази.		Ставицька	
не бздо в канстру	не бздо в канстру			Уживається для вираження заклику перестати боятися чогось.		Ставицька	
бздун	бздун			1. Той, хто псує повітря, поширює сморід (перев. про старих чоловіків). 2. Боягуз.		Бобровник	Ставицька
бздун	бздуняка			1. Той, хто псує повітря, поширює сморід (перев. про старих чоловіків). 2. Боягуз.		Бобровник	Ставицька
бздун	бздур			1. Той, хто псує повітря, поширює сморід (перев. про старих чоловіків). 2. Боягуз.		Бобровник	Ставицька

Рисунок 3.8. Таблиця в Google Sheets з лексикографічною інформацією

Для створення вебінтерфейсу словника лексикографічні дані у вигляді електронної таблиці треба було перетворити у базу даних для опрацювання комп'ютером. На цьому етапі було з'ясовано, що укладена таблиця не відповідає правилам нормалізації баз даних (Greenhouse, 2024). Нормалізація – це метод проєктування бази даних, який дозволяє привести базу даних до мінімальної надмірности; процес видалення надлишкових даних (Greenhouse, 2024). Надмірність даних створює передумови для появи різних аномалій, знижує продуктивність та погіршує гнучкість керування базою даних (Greenhouse, 2024). Для усунення цієї проблеми було вирішено розбити електронну таблицю на декілька менших таблиць, поєднаних між собою зв'язками.

Зрештою на основі однієї загальної таблиці було створено п’ять нових таблиць.

1) Таблиця **Слово** складалася з п’яти колонок: *ID (Primary Key)*, де вказувався ідентифікаційний номер реєстрового слова, *Слово* – колонка для запису самої лексеми, *Сфера_вживання* – назва сфери, з якої походить дане слово, *Відповідник* – український відповідник лексеми та *Тлумачення*, до якої вносились дефініції конкретної лексеми із колонки *Слово*. Колонки *Сфера_вживання* та *Відповідник* були наповнені лише для деяких сленгізмів. Усього таблиця містила 1043 рядки. Це означає, що «Словник обсценізмів та сленгу» містить 1043 статті, оскільки статті писалися за реєстровим словом, а не за кожним варіянтом окремо.
2) Таблиця **Джерело** складалася з двох колонок: *ID* – унікального ідентифікаційного номера джерела та *Назва*, де зазначалася назва відповідного лексикографічного джерела. Ця таблиця була створена для уникнення дублювання назв джерел у базі даних. Усього таблиця містила три записи, що відповідали трьом джерелам, використаним під час укладання словника.
3) Таблиця **Слово_джерело** складалася з двох колонок: *ID_слова* та *ID_джерела*. Її призначенням було реалізувати зв’язок між таблицями *Слово* та *Джерело*. Оскільки одне слово могло бути зафіксоване у двох джерелах, а кожне джерело містило значну кількість слів, між цими таблицями існував зв’язок типу «багато до багатьох».
4) Таблиця **Варіянт** складалася з трьох колонок: *ID_варіянту* – унікального ідентифікаційного номера варіянту, *ID_слова* – зовнішнього ключа, який був пов’язаний із записом відповідної лексеми у таблиці *Слово*, та *Варіянт*, до якої вносилися всі зафіксовані форми лексеми, включаючи їхні фонетичні, орфографічні та словотвірні варіянти. Усього таблиця налічувала 1790 записів.
5) Таблиця **Приклад** складалася з чотирьох колонок: *ID_прикладу* – унікального ідентифікаційного номера прикладу, *ID_слова*

– зовнішнього ключа, що пов’язував приклад із відповідною словниковою статтею, *Приклад*, до якої вносився приклад-ілюстрація вживання лексеми, та *Автор_прикладу*, де зазначався нік автора у соцмережі X (колшній Twitter). Усього до таблиці було внесено 33 приклади-ілюстрації.

Після створення таблиць у Google Sheets їх було конвертовано в бази даних SQLite. Для цього кожну таблицю було збережено у форматі TSV (Tab-Separated Values), який забезпечує коректне перенесення табличних даних між різними програмними середовищами та підтримує автоматизоване опрацювання інформації. Використання цього формату дозволило уникнути втрати структури даних під час експорту та спростило їх подальше завантаження до бази даних.

Для створення структури бази даних було використано програму SQLite Expert. SQLite Expert є спеціялізованим середовищем для проєктування, адміністрування та редагування баз даних SQLite, яке надає засоби для створення таблиць, визначення первинних і зовнішніх ключів, налаштування індексів та виконання SQL-запитів (SQLite Expert, 2025). Однією з переваг цього інструмента є підтримка візуального проєктування структури бази даних, що значно полегшує розробку складних схем із великою кількістю взаємопов’язаних таблиць.

Після створення структури бази даних залишилося імпортувати лексикографічні дані з TSV-файлів до відповідних таблиць SQLite. Оскільки проведення цього етапу вручну потребувало великої кількости часу, було розроблено програмний код мовою Python, який автоматично вкионував імпорт даних. Програмний код забезпечив однаковий формат запису даних у всіх таблицях та мінімізував ризик виникнення технічних помилок під час наповнення бази даних.

Під час проєктування бази даних особливу увагу було приділено встановленню зв’язків між таблицями за допомогою зовнішніх ключів. Зовнішній ключ є механізмом, що забезпечує цілісність даних шляхом контролю відповідности записів між пов’язаними таблицями. Завдяки використанню зовнішніх ключів кожен запис у таблицях «Варіянт», «Приклад» та

«Слово_джерело» міг бути пов’язаний лише з наявним записом у таблиці «Слово», що унеможливлювало появу так званих «осиротілих» записів і підтримувало логічну узгодженість бази даних.

У результаті було створено реляційну базу даних, що складалася з п’яти взаємопов’язаних таблиць. Така структура забезпечила ефективне зберігання лексикографічної інформації, зменшення надмірности даних і можливість швидкого виконання пошуку за різними параметрами. Створена база даних стала основою для подальшої розробки вебінтерфейсу словника та реалізації функцій пошуку, фільтрації й відображення словникових статей.

3.3. Вебінтерфейс словника обсценної лексики та сленгу

Інтерфейс словника було вирішено створити у вигляді вебсторінки, оскільки це простий і зрозумілий формат для доступу до бази даних. Вебінтерфейс було створено за допомогою мови програмування PHP та мов розмітки HTML та CSS. У ході розробки будь-якого проєкту виникає питання вибору мови програмування. Для вебзастосунків найпоширенішими є PHP та JavaScript (Zend Technologies, 2023).

PHP – (рекурсивний акронім від PHP: Hypertext Preprocessor) — це широко використовувана мова сценаріїв загального призначення з відкритим кодом, яка особливо підходить для веброзробки та може бути вбудована в HTML (PHP Group, 2025). На відміну від клієнтських мов (напр. JavaScript), які працюють у браузері користувача, PHP працює на веб-сервері та обробляє код, перш ніж сторінка надсилається користувачеві.(Fasthosts, 2024) Це робить його ідеальним для створення функцій, які залежать від зберігання, отримання або оновлення даних за лаштунками (Fasthosts, 2024). PHP використовується у бекенд-розробці, тобто розробці серверної частини сайту, невидимої для користувача (WEZOM, 2023). Код, написаний PHP, відповідає за логіку обробки запитів, збереження та отримання даних, управління обліковими записами, авторизацію, взаємодію з зовнішніми API тощо (WEZOM, 2023).

PHP дозволяє виконувати широкий спектр реальних завдань, зокрема:

- створення динамічних веб-сайтів на основі баз даних;
- обробка форм та введення даних користувачем;
- створення сторінок, що відображаються на сервері;
- керування логінами, сеансами та обліковими записами користувачів (Fasthosts, 2024).

Додатковою перевагою PHP є те, що вона – проста для вивчення, але пропонує багато розширених функцій для професійного програміста (PHP Group, 2025). За словами розробників, “з PHP майже кожен може розпочати роботу та писати прості скрипти в найкоротші терміни” (PHP Group, 2025).

JavaScript – це мова програмування, яка використовується як на боці сервера, так і на боці фронтенду (WEZOM, 2023). У поєднанні з мовами розмітки HTML та CSS вона забезпечує вебсторінкам динамічність та інтерактивність, створюючи анімації (WEZOM, 2023).

За даними Statista, станом на 2024 рік 69% розробників у всьому світі використовують JavaScript, а ще 5% планують перейти на цю мову (Statista, 2024). Звіт показує, що станом на кінець 2019 року це була найпопулярніша мова програмування у світі.

JavaScript є клієнтською мовою програмування (Zend Technologies, 2023). Вона працює в браузерах, і завдяки її широкій популярності більшість браузерів можуть підтримувати JavaScript. Кожен браузер має JavaScript Engine (“рушій”, інтерпретатор коду, написаного мовою програмування JavaScript) для виконання коду (Code Institute, 2024). До таких інтерпретаторів належать SpiderMonkey та V8 (Code Institute, 2024).

Крім коду, який запускається у браузерах, розробники можуть створювати серверний JavaScript-код за допомогою спеціяльного середовища для виконання коду JavaScript – Node.js (Node.js Foundation, 2025). Вебринок також пропонує різноманітні JavaScript-фреймворки, такі як AngularJS, ReactJS, NodeJS тощо(Code Institute, 2024). Фреймворки – це набір інструментів, бібліотек та правил, який використовується для створення програмних додатків (Brainlab,

2023). Ці фреймворки значно скорочують час і зусилля розробки вебсайтів і додатків на основі JavaScript (Code Institute, 2024).

Оскільки основною задачею PHP є реалізація серверної складової роботи вебзастосунку, а за допомогою JavaScript можна розробляти динамічний інтерфейс, дві мови програмування можна поєднувати в межах одного проєкту. Проте вивчення двох мов замість однієї є часозатратним, тому для розробки дипломного проєкту словника було вирішено обрати лише одну.

Попри такі переваги JavaScript як створення динамічного інтерфейсу, можливість створення як серверної так і клієнтської частини коду, велика кількість спеціяльних фреймворків для виконання різних задач тощо, для розроблення вебзастосунку було обрано мову PHP.

Хоча JavaScript є надзвичайно популярною в розробці вебпроєктів, у конкретному випадку із розробкою студентського вебсайту словника переваги PHP є більш вагомими. Насамперед PHP є відносно простою для вивчення мовою, що дозволяє швидко створювати функціональні вебзастосунки навіть за відсутности значного досвіду веброзробки. Крім того, PHP має вбудовані засоби для роботи з базами даних, завдяки чому може безпосередньо взаємодіяти з базою SQLite без використання додаткових середовищ виконання чи програмних платформ. На відміну від JavaScript, для реалізації серверної логіки якого зазвичай необхідно використовувати середовище Node.js, PHP виконується безпосередньо на вебсервері та забезпечує просте отримання, обробку й відображення даних. Це зробило PHP оптимальним вибором для створення вебінтерфейсу словника та реалізації пошуку за лексикографічною базою даних.

Для ефективної розробки та швидкого тестування вебзастосунку було використано локальний вебсервер XAMPP. XAMPP – це простий у встановленні та налаштуванні пакет для розгротання локального вебсерверу на власному комп’ютері. Він є кросплатформним, підтримується на Windows, Linux та macOs, містить Apache, MySQL (MariaDB), PHP та Perl (ці пакети програмного забезпечення використовуються для запуску динамічних вебсайтів або серверів) (HostPro, 2024). Використання локального вебсерверу значно спростило роботу

із графічним інтерфейсом словника та прискорило процес виявлення та виправлення помилок роботи коду.

У результаті виконаної роботи було створено вебсторінку «Словника обсценізмів та сленгу», яка забезпечує доступ до лексикографічної бази даних через вебінтерфейс. Для спрощення навігації вебзастосунок було поділено на чотири основні вкладки меню: *Головна, Реєстр словника, Пошук* та *Джерела*.

Головна сторінка виконує ознайомчу функцію. На ній розміщено коротку інформацію про мету створення словника, його структуру та особливості. Також ця вкладка містить стислий опис поняття обсценної лексики. На сторінці подається інформація про джерельну базу словника та посилання на окремий розділ «Джерела», де подано детальніші відомості про використані лексикографічні праці.

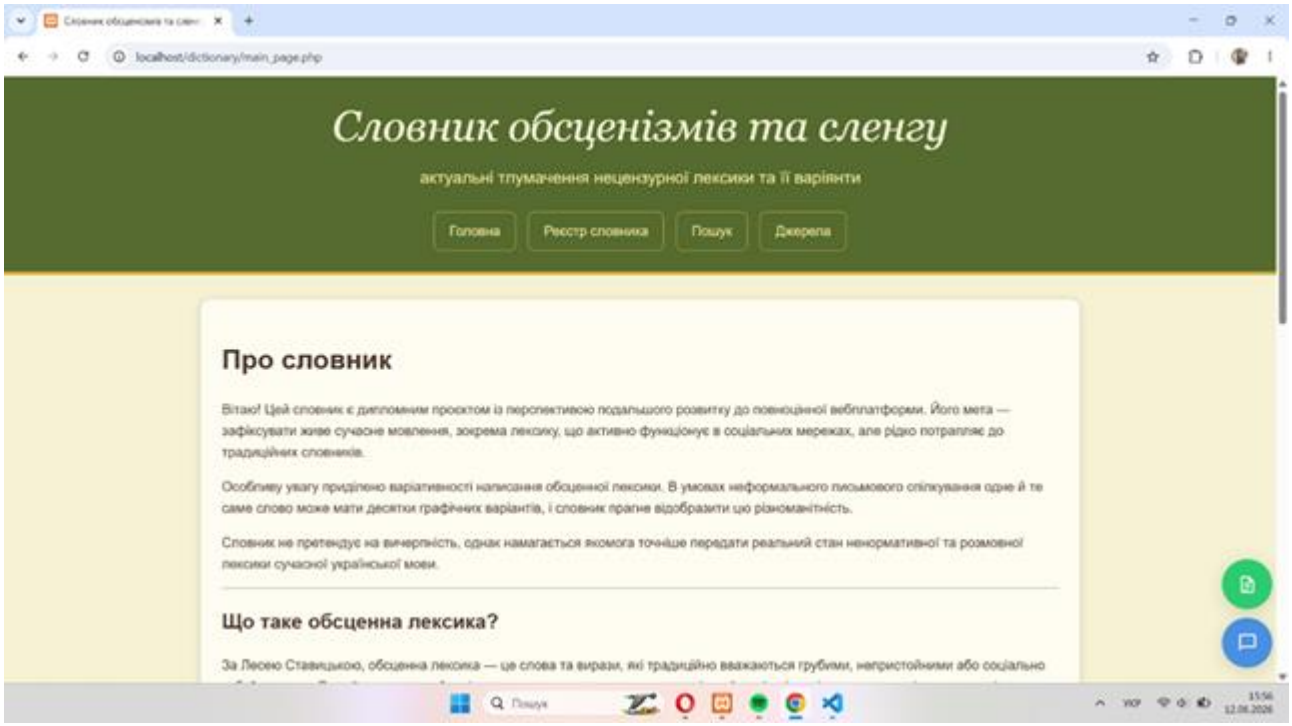


Рисунок 3.9. Головна сторінка вебсайту словника

Вкладка **Реєстр словника** призначена для перегляду повного словникового фонду у вигляді переліку лексем. Список упорядкований за алфавітом і включає як реєстрові слова, так і їхні варіянти. Для зручности

сприйняття великий обсяг даних організовано у три колонки. Кожна лексема є гіперпосиланням, за допомогою якого користувач може перейти безпосередньо до відповідної словникової статті. Такий підхід забезпечує швидкий доступ до необхідної інформації навіть без використання пошукової системи.

Основною функціональною складовою вебзастосунку є вкладка **Пошук**. Вона містить текстове поле для введення пошукового запиту та кнопку запуску пошуку. Механізм пошуку працює не лише з реєстровими словами, а й з усіма зафіксованими у базі даних варіантами лексем. Так, наприклад, пошук за словом *хєр* дозволяє знайти статтю до реєстрової форми *хер*, якщо відповідний варіянт зафіксовано у базі даних.

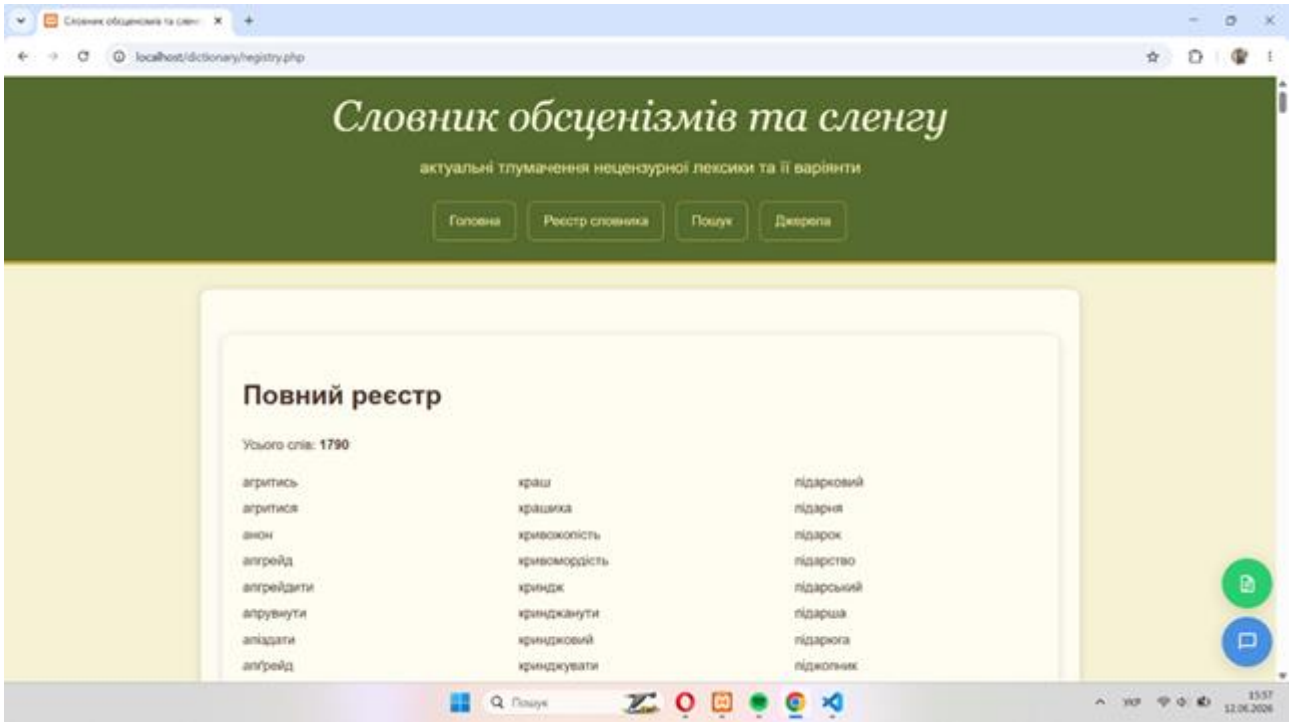


Рисунок 3.10. Сторінка із повним реєстром словника

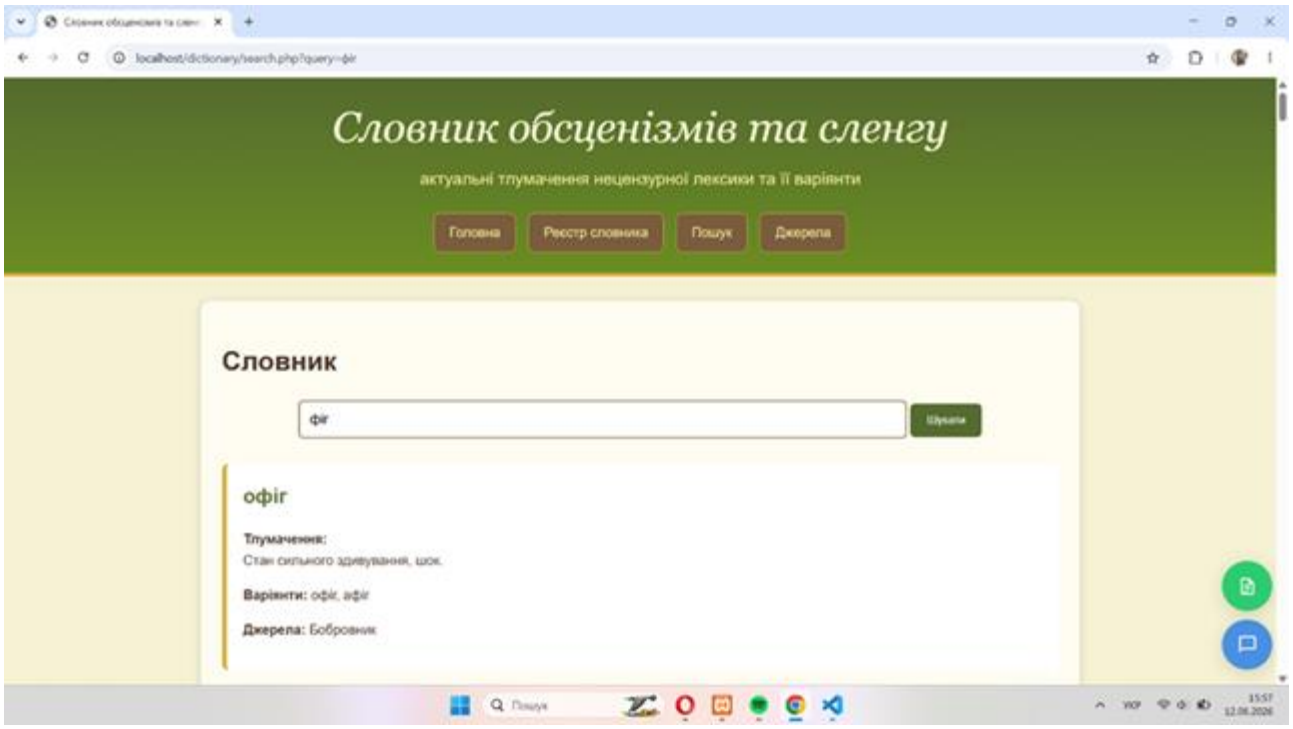


Рисунок 3.11. Сторінка із пошуком за словником (пошук слова фір*)*

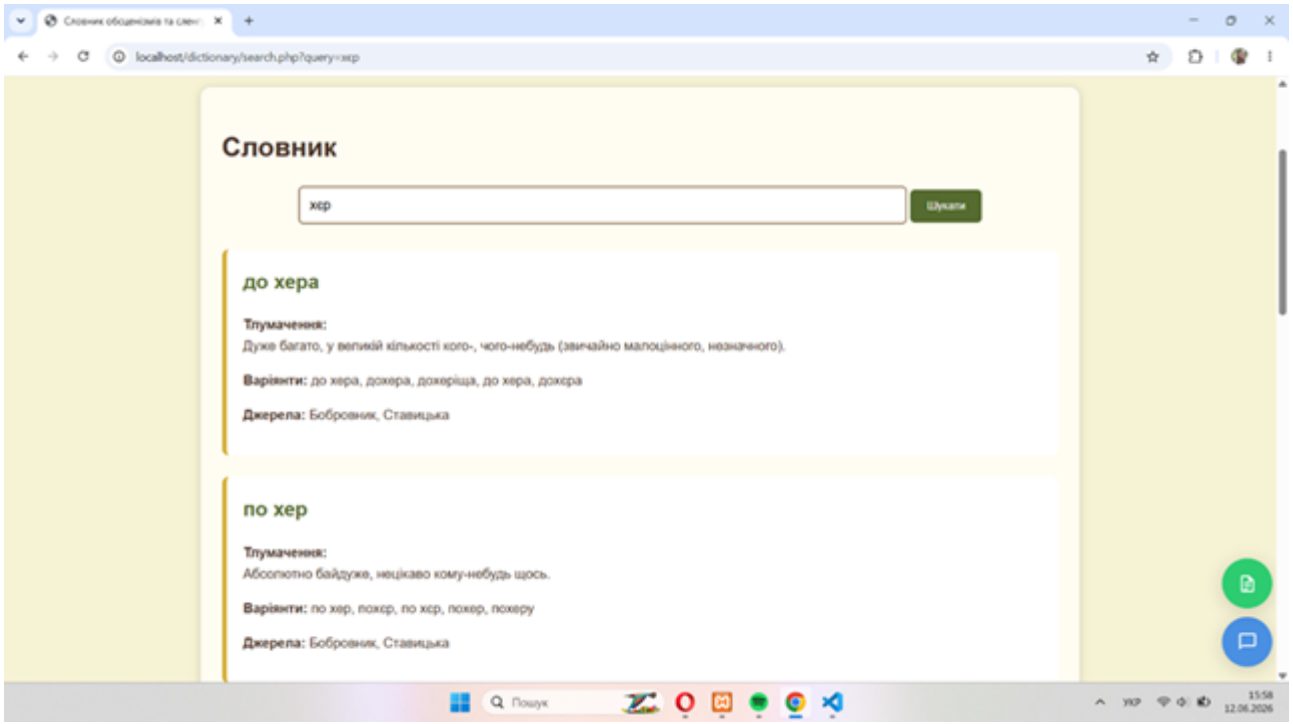


Рисунок 3.12. Сторінка із пошуком за словником (пошук слова хєр)

Ще однією особливістю пошукової системи є підтримка часткового збігу символів. Пошук виконується за принципом входження пошукового фрагмента до тексту реєстрового слова або його варіянтів. Це означає, що користувачеві не

обов’язково вводити повну лексему. Наприклад, введення послідовности символів *фіг* дозволить отримати результати для таких слів, як *фіговий*, *фігачити*, *до фіга* та інших одиниць, у складі яких міститься відповідний фрагмент.

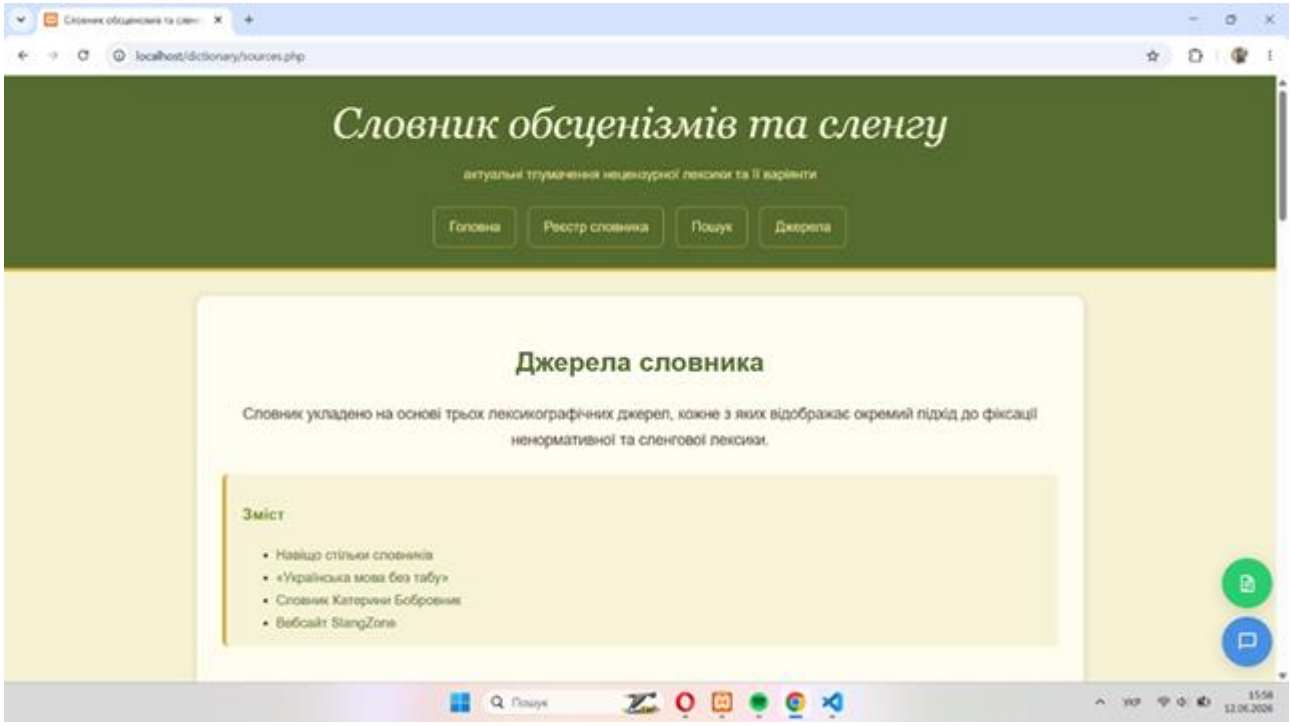


Рисунок 3.13. Сторінка Джерела *(ч. 1)*

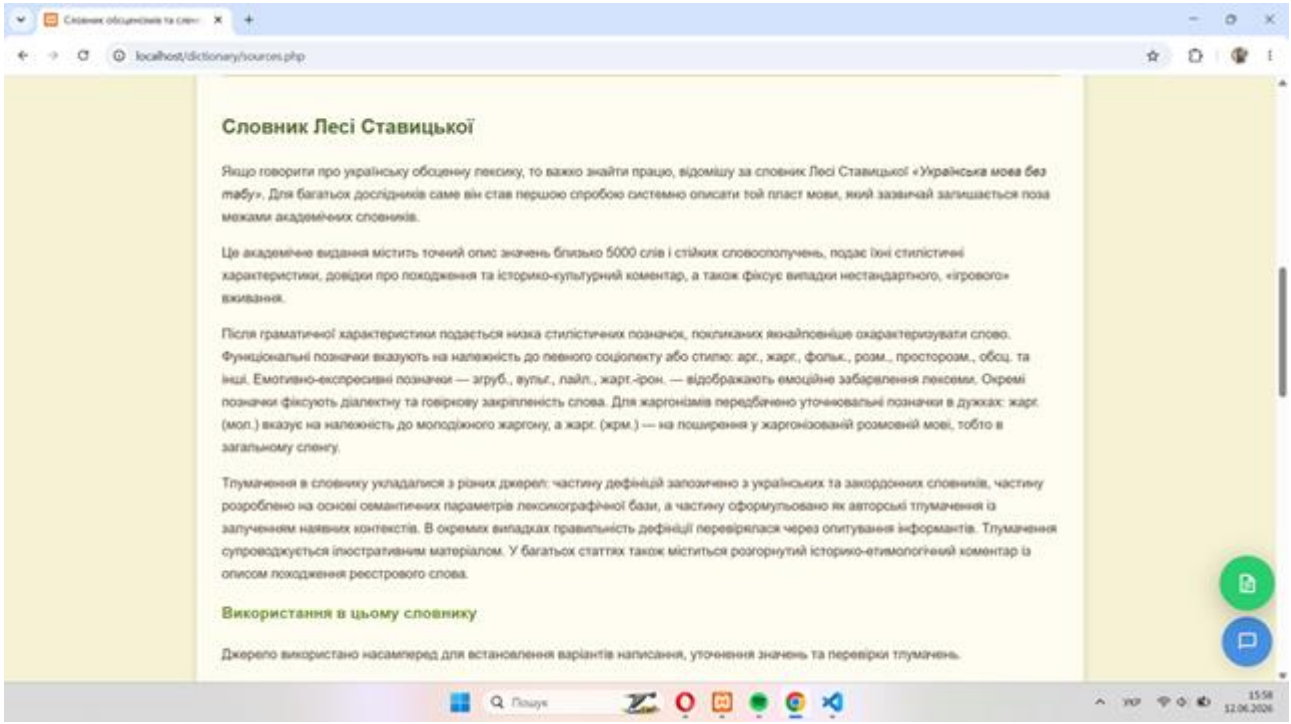


Рисунок 3.14. Сторінка Джерела *(ч. 2)*

Вкладка **Джерела** містить докладну інформацію про лексикографічні праці, що стали основою для укладання словника. У цьому розділі подано відомості про працю Лесі Ставицької «Українська мова без табу. Словник нецензурної лексики та її відповідників. Обсценізми, евфемізми, сексуалізми», словник Катерини Бобровник, а також онлайн-словник нецензурної лексики SlangZone.

Висновки до третього розділу

Після опрацювання словників К. Бобровник та Л. Ставицької, а також вмісту онлайн-словника SlangZone до «Словника обсценізмів та сленгу» було відібрано загалом 1790 лексем. Відібрані матеріяли було розподілено у групи за варіянтами. На вибір реєстрового слова впливало стилістичне забарвлення лексеми, наявність пестливих суфіксів чи суфіксів згрубілости, а також частота вживання у соціяльній мережі X. Таким чином було визначено 1043 реєстрові слова, які стали основою словникових статей.

У майбутньому перспективним напрямком є ще більше розширення реєстрового списку словника за допомогою програмної генерації фонетичних варіянтів. Це можна реалізувати через виявлення закономірних фонетичних варіювань деяких морфем (напр., для всіх лексеми з коренем *-єб-* створити аналогічні форми, де даний корінь змінюється на *-їб-, - iп-, -йоб-, -йоп-, -їп-* тощо). Такий підхід перетворить "Словник обсценізмів та сленгу" на універсальне джерело довідки для широкого кола користувачів.

На основі трьох лексикографічних джерел було створено реляційну базу даних словника. Вона складалася з п'яти взаємопов'язаних таблиць та містила інформацію про реєстрові слова, їхні варіянти, джерела фіксації, сферу вживання, тлумачення й приклади-ілюстрації. Для забезпечення ефективного зберігання та опрацювання лексикографічних даних структура бази даних була нормалізована, що дозволило мінімізувати надмірність інформації.

На наступному етапі було розроблено вебсторінку словника обсценізмів та сленгу, яка забезпечує доступ до бази даних через користувацький інтерфейс. Створений вебзастосунок поєднує можливості перегляду повного реєстру лексем, швидкого пошуку за реєстровими словами та їхніми варіянтами, а також ознайомлення з джерельною базою словника.

ВИСНОВКИ

У кваліфікаційній роботі було досліджено особливості лексикографічного опрацювання української обсценної лексики та сленгу, а також розроблено електронний словник, призначений для збирання, систематизації та представлення відповідного мовного матеріялу. Актуальність роботи зумовлена зростанням ролі неформальної комунікації в сучасному українському мовному просторі та потребою у створенні цифрових лексикографічних ресурсів, які б фіксували й описували такі мовні одиниці.

У теоретичній частині роботи було розглянуто поняття обсценної лексики та сленгу, проаналізовано основні підходи до їхнього визначення й класифікації в сучасному мовознавстві. Окрему увагу приділено проблемам лексикографічного опису стилістично маркованої лексики, а також особливостям укладання електронних словників. Проведений аналіз засвідчив, що обсценізми та сленгізми становлять важливу складову сучасної мовної практики та потребують системного документування і впорядкування.

У практичній частині роботи було сформовано джерельну базу словника на основі трьох лексикографічних ресурсів: словника Лесі Ставицької «Українська мова без табу. Словник нецензурної лексики та її відповідників. Обсценізми, евфемізми, сексуалізми», словника Катерини Бобровник та онлайн-словника SlangZone. У результаті опрацювання цих джерел було відібрано 1790 лексичних одиниць. Після аналізу варіянтности та встановлення реєстрових форм було сформовано 1043 словникові статті. Під час відбору реєстрових слів враховувалися стилістичні характеристики лексем, наявність словотвірних модифікацій та частота вживання окремих форм у сучасному мовленні.

Для забезпечення ефективного зберігання та опрацювання лексикографічних даних було спроєктовано реляційну базу даних. У процесі її створення здійснено нормалізацію первинної таблиці, що дозволило усунути надмірність даних і забезпечити цілісність інформації. Створена база даних складалася з п’яти взаємопов’язаних таблиць і містила відомості про реєстрові слова, їхні варіянти, джерела фіксації, тлумачення, сфери вживання та приклади-

ілюстрації. Для реалізації бази даних було використано систему SQLite, а процес імпорту даних автоматизовано за допомогою програмного коду мовою Python.

На основі створеної бази даних було розроблено вебзастосунок «Словника обсценізмів та сленгу». Для реалізації серверної частини було використано мову програмування PHP, що забезпечила безпосередню взаємодію з базою даних та обробку користувацьких запитів. Вебінтерфейс містить сторінки з інформацією про словник, повний реєстр лексем, пошукову систему та відомості про використані джерела. Реалізований механізм пошуку дозволяє знаходити словникові статті як за реєстровими словами, так і за їхніми варіянтами, що значно підвищує зручність користування ресурсом.

Таким чином, поставлену мету роботи було досягнуто, а всі визначені завдання виконано. Результатом дослідження став функціональний електронний словник української обсценної лексики та сленгу, який поєднує лексикографічну систематизацію мовного матеріялу із сучасними засобами його цифрового представлення. Створений ресурс може бути використаний для подальших лінгвістичних досліджень, вивчення сучасної української розмовної мови, а також як основа для розширення й удосконалення електронних лексикографічних систем. Перспективою подальшої роботи є доповнення словника новими лексичними одиницями, розширення корпусу прикладів уживання, удосконалення пошукових механізмів та інтеграція додаткових лінгвістичних параметрів для поглибленого аналізу мовного матеріялу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бобровник, К. А. (2019) “Автоматичне визначення токсичних коментарів (на матеріалі соціальної мережі Twitter)”.
2. Парфіненко, А. О. (2013) “Мовна варіативність у фокусі сучасних лінгвістичних досліджень (на матеріалі праць британських та американських вчених)”, Київ: Studia Linguistica, 7.
3. Ставицька, Л. О. (2003) “Короткий словник жарґонної лексики української мови = A Short Dictionary of Ukrainian Slang : містить понад 3200 слів і 650 стійких словосполучень”, Київ: Критика.
4. Ставицька, Л. О. (2008) “Українська мова без табу. Словник нецензурної лексики та її відповідників”, Київ: Критика.
5. Berne, E. (1970), “Sex in human loving”.
6. Chomsky, Noam (1957), Syntactic Structures, The Hague/Paris: Mouton, ISBN 978-3-11-021832-9
7. Fishman, Joshua A. (1965) “Who Speaks What Language to Whom and When?” La Linguistique 1, no. 2: 67–88. URL: http://www.jstor.org/stable/30248773 (дата звернення: 26.05.2026).
8. Jurafsky, Daniel & Martin, James. (2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.
9. Labov, William. (1963) “The social motivation of a sound change.” WORD 19: 273-309.
10. Robinson, J. A. (2010) ‘Awesome insights into semantic variation.’ In Geeraerts, D., Kristiansen, G. & Piersman, Y. (eds), Advances in Cognitive Sociolinguistics. Berlin: Mouton de Gruyter, pp. 85–110.Google Scholar
11. Bitdefender. (2023) ‘Twitter thwarts third-party data scraping by rolling out new sign-in feature to view tweets and profiles’. Available at: Bitdefender Blog. URL: https://www.bitdefender.com/en-us/blog/hotforsecurity/twitter-thwarts-third-

party-data-scraping-by-rolling-out-new-sign-in-feature-to-view-tweets-and-profiles (дата звернення: 26.05.2026).

12. Цигвінцева Ю. О. (2025) «Лексикографічне моделювання неосемантизмів у словниках української мови першої чверті XXI століття», Київ: Мовознавство. URL: https://ukr.movoznavstvo.knu.ua/uk/article/view/3806 (дата звернення: 27.05.2026).
13. Цигвінцева Ю. О. (2020) «Неосемантизми в складі нових українських фразеологізмів», Ужгород: Науковий вісник Ужгородського університету. Серія Філологія.
14. Білецька О. (2025) «Мовні інновації війни: українські військові неосемантизми та їх переклад німецькою мовою», Донецьк: *Лінгвістичні студії*.
15. Кислюк Л. П. (2025) «Нова лексика воєнного часу та її лексикографування», Київ: Лексикографічний бюлетень, 33, с. 33–48.
16. Клименко Н. Ф, Карпіловська Є. А, Кислюк Л. П. (2008) «Динамічні процеси в українському лексиконі», Київ: Вид. Дім Дмитра Бураго.
17. Цигвінцева Ю. О. (2024) «Неосемантизація в сучасній українській мові: чинники активізації внутрішніх ресурсів номінації», Вінниця: Нілан-ЛТД.
18. Geeraerts Dirk, Speelman Dirk, Heylen Kris, Montes Mariana, De Pascale, Stefano Franco, Karlien Lang (2024) *Lexical Variation and Change*, Oxford University Press. URL: https://www.researchgate.net/publication/375860116_Lexical_Variation_and_Change_A_Distributional_Semantic_Approach (дата звернення: 25.05.2026).
19. Daphna Keidar, Andreas Opedal, Zhijing Jin, and Mrinmaya Sachan (2022), Slangvolution: A Causal Analysis of Semantic Change and Frequency Dynamics in Slang. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. URL: Slangvolution: A Causal Analysis of Semantic Change and Frequency Dynamics in Slang – ACL Anthology (дата звернення: 26.05.2026).

20. Барсукова О. (2022) «Слово “блєдіна”, яким називають російські ракети, внесли до словника Urban Dictionary», Київ: Українська правда. Життя. URL: https://www.researchgate.net/publication/375860116_Lexical_Variation_and_Change_A_Distributional_Semantic_Approach (дата звернення: 25.05.2026).
21. Шевченко В. (2022) «Що таке «блєдіна»: від марки дитячого харчування до символу руснявої ракети», Київ: Na Chasi. URL: https://nachasi.com/creative/2022/06/22/what-is-bledina/ (дата звернення: 24.05.2026).
22. Urban Dictionary. (2024) ‘Bledina’. Available at: Urban Dictionary. URL: https://www.urbandictionary.com/define.php?term=bledina (дата звернення: 24.05.2026).
23. Должикова Т. (2013) «Сленгізми в сучасному українському художньому тексті», *Волинь філологічна: текст і контекст*, (15), 92-100.
24. Greenhouse (2024) «Нормалізація баз даних», URL: Нормалізація баз даних. – Greenhouse (дата звернення: 26.05.2026).
25. SQLite Expert. (2025) *SQLite Expert Professional Documentation*. Available at: SQLite Expert. URL: https://www.sqliteexpert.com/index.html (дата звернення: 26.05.2026).
26. PHP Group. (2025) *PHP Manual: Introduction*. Available at: PHP Documentation. URL: https://www.php.net/manual/en/introduction.php (дата звернення: 26.05.2026).
27. Zend Technologies. (2023) ‘PHP vs JavaScript: Key Differences and Use Cases’. Available at: Zend Blog. URL: https://www.zend.com/blog/php-vs-javascript (дата звернення: 24.05.2026).
28. Fasthosts. (2024) ‘PHP vs JavaScript: Which Language Is Better?’. Available at: Fasthosts Blog. URL: https://www.fasthosts.co.uk/blog/php-vs-javascript/ (дата звернення: 25.05.2026).

29. WEZOM. (2023) «Що таке Back-End розробка», Харків: WEZOM Blog. URL: https://wezom.com.ua/ua/blog/chto-takoe-back-end-razrabotka (дата звернення: 25.05.2026).
30. freeCodeCamp. (2023) 'PHP vs JavaScript: Which Technology Will Suit Your Business Better?'. Available at: freeCodeCamp. URL: https://www.freecodecamp.org/news/php-vs-javascript-which-technology-will-suit-your-business-better/ (дата звернення: 25.05.2026).
31. Statista. (2024) 'Worldwide Software Developer Survey: Programming Languages Used'. Hamburg: Statista. URL: https://www.statista.com/statistics/869092/worldwide-software-developer-survey-languages-used/ (дата звернення: 25.05.2026).
32. Code Institute. (2024) 'PHP vs JavaScript: Which Should You Learn?'. Available at: Code Institute Blog. URL: https://codeinstitute.net/global/blog/php-vs-javascript/ (дата звернення: 25.05.2026).
33. Node.js Foundation. (2025) «Про Node.js». Available at: Node.js Documentation. URL: https://nodejs.org/uk/about (дата звернення: 25.05.2026).
34. Brainlab. (2023) «Що таке фреймворк: пояснюємо простими словами», Київ: Brainlab. URL: https://brainlab.com.ua/uk/blog-uk/shho-take-frejmvork-poyasnyuyemo-prostymy-slovamy (дата звернення: 26.05.2026).
35. HostPro. (2024) «Встановлення та налаштування локального вебсервера XAMPP на Windows», Київ: HostPro Wiki. URL: https://hostpro.ua/wiki/ua/instructions/installing-and-configuring-the-local-xampp-web-server-on-windows/ (дата звернення: 25.05.2026).
36. Ljung M. (2011) *Swearing: A Cross-Cultural Linguistic study*. Sweden: University of Stockholm. URL: https://www.academia.edu/91676772/Swearing_A_Cross_Cultural_Linguistic_Study (дата звернення: 23.05.2026).

37. Shiri Lev-Ari & Ryan McKay (2022) ‘The sound of swearing: Are there universal patterns in profanity?’, *Psychonomic Bulletin & Review*, 30, pp. 1–22. URL: https://link.springer.com/article/10.3758/s13423-022-02202-0 (дата звернення: 23.05.2026).
38. Bergen, B. K. (2016) *What the F: What Swearing Reveals About Our Language, Our Brains, and Ourselves*. New York: Basic Books. URL: https://books.google.com.ua/books?hl=uk&lr=&id=WqtVDgAAQBAJ&oi=fnd&pg=PT7&dq=swearing&ots=gz7i6_BDH1&sig=b5xp9PtleveqIs54aaFtTVnieMQ&redir_esc=y#v=onepage&q=swearing&f=false (дата звернення: 23.05.2026).
39. Montagu, A. (1967) *The Anatomy of Swearing*. New York: Macmillan. URL: https://archive.org/details/anatomyofswearin00mont (дата звернення: 25.05.2026).
40. Podolyak, M. (2022) ‘Post on X (formerly Twitter), 28 March 2022’. URL: https://x.com/MelaniePodolyak/status/1508746802048417795?s=20 (дата звернення: 27.05.2026).
41. eighteen_th. (2026) ‘Post on X (formerly Twitter), 13 June 2026’. URL: https://x.com/eighteen_th/status/2065671503589360120?s=20. (дата звернення: 27.05.2026).
42. Словник SlangZone. URL: https://www.slangzone.net/uk/ (дата звернення: 15.05.2026).

ДОДАТКИ

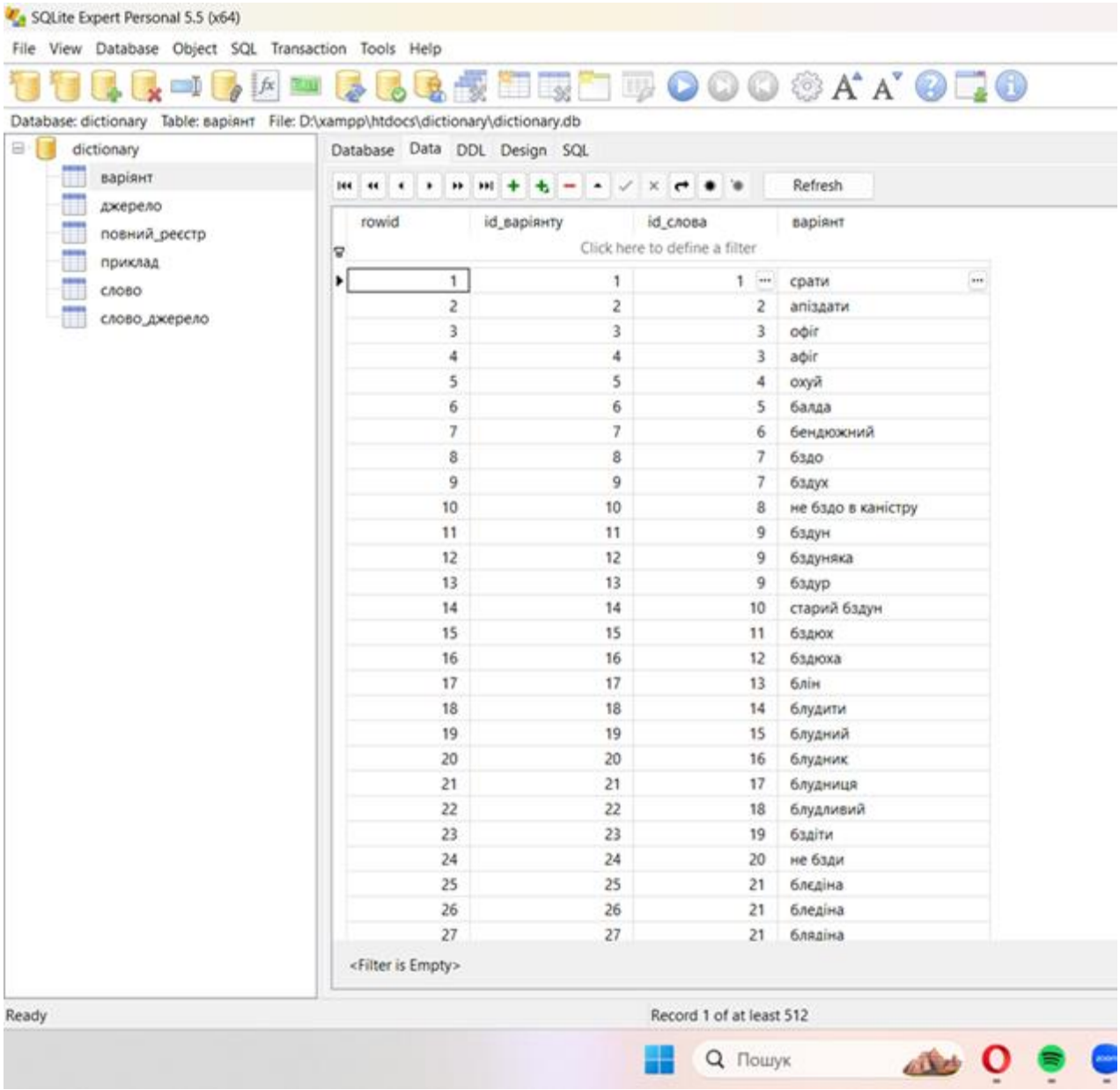
Додаток А. Відео роботи вебсайту

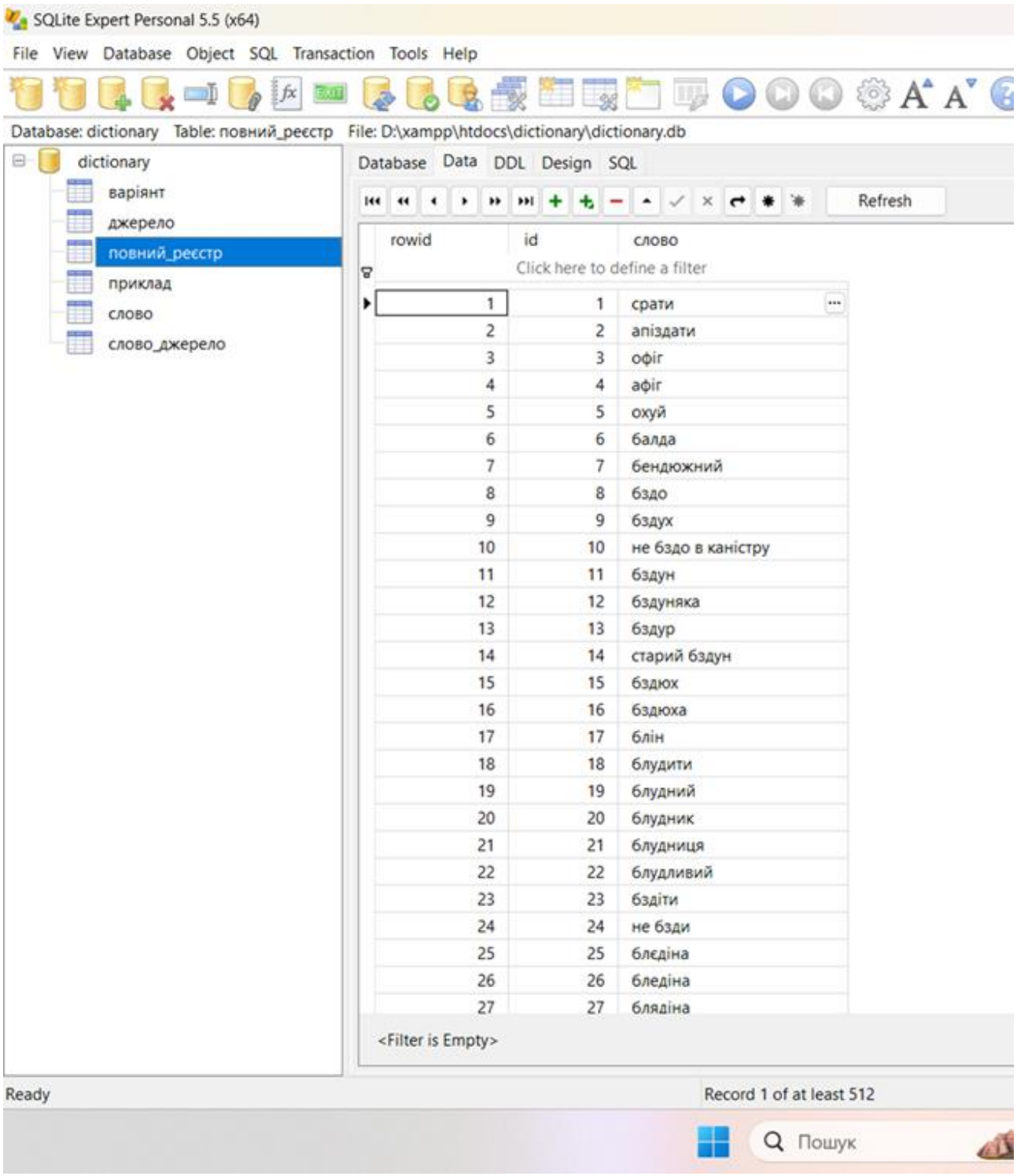
Відео доступне до перегляду за посиланням.

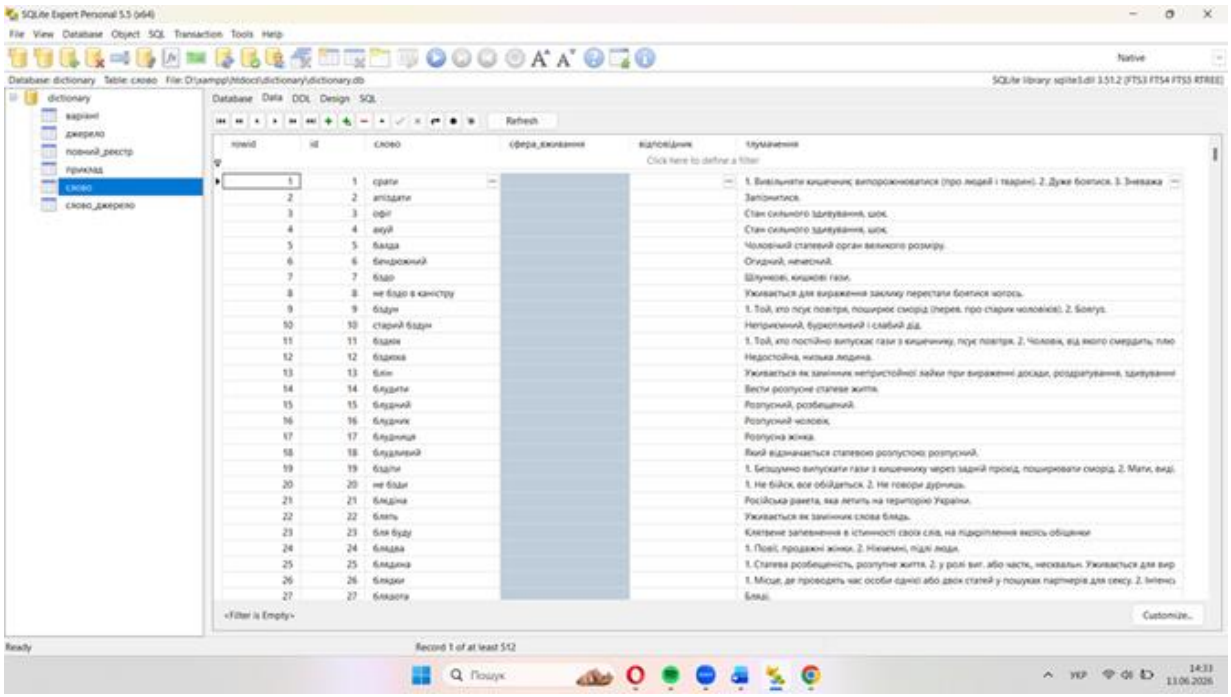
Додаток Б. Програмне забезпечення продукту

Папка із усіма програмами для підтримки роботи сторінок вебсайту доступна до перегляду за посиланням.

Додаток В. Вигляд бази даних із лексикографічної інформацією




SQLite Expert Personal 5.5 (x64)
File View Database Object SQL Transaction Tools Help
Database: dictionary Table: повний_реєстр File: D:\xampp\htdocs\dictionary\dictionary.db
dictionary
варіянт
джерело
повний_реєстр
приклад
слово
слово_джерело
Database Data DDL Design SQL
Refresh
rowid id слово
Click here to define a filter
1 1 срати
2 2 апіздати
3 3 офіг
4 4 афіг
5 5 охуй
6 6 балда
7 7 бендюжний
8 8 бздо
9 9 бздух
10 10 не бздо в каністру
11 11 бздун
12 12 бздуняка
13 13 бздур
14 14 старий бздун
15 15 бздюх
16 16 бздюха
17 17 блін
18 18 блудити
19 19 блудний
20 20 блудник
21 21 блудниця
22 22 блудливий
23 23 бздіти
24 24 не бзди
25 25 блєдіна
26 26 бледіна
27 27 блядіна
<Filter is Empty>
Ready
Record 1 of at least 512
Пошук



Додаток Г. База даних із лексикографічною інформацією

База даних, на основі якої працює вебсайт, доступна до перегляду за посиланням.