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

Створення чат-бота "Повертаємо репресовану мову" на платформі месенджера Telegram

Кваліфікаційна робота бакалавра
студентки 4 курсу
освітньої програми
«Прикладна (комп'ютерна) лінгвістика та англійська мова»
спеціальності – 035.10 Філологія (прикладна лінгвістика)
галузі знань – 03 гуманітарні науки
Анастасії КОСЕНКО
Науковий керівник:
к. філол. н, доцент **Оксана ЗУБАНЬ**

«Допущено до захисту»
Протокол засідання
кафедри української мови та прикладної лінгвістики
протокол № 15 від «06» 06 2024 року

завідувач кафедри _____(підпис)
к.філол.н., доц. Сергій РІЗНИК

КИЇВ – 2024

ЗМІСТ

ВСТУП

Актуальність дослідження зумовлено збільшенням наукового та практичного інтересу до аспектів репресованої лексики, що постають перед сучасними мовознавцями. Лінгвістика є наукою, яка вивчає мову як конструкт, що постійно розвивається та зазнає змін. При цьому трансформації можуть зазнавати не лише лексичний склад мови, але й інші аспекти мовознавства: граматика, семантика, фонетика, синтаксис, структура та правопис. Лінгвістика також акцентує увагу на зв'язок мови зі змінами, що відбуваються у суспільстві. При цьому мова розглядається і як складник ідентичності особистості та спосіб національного самовизначення та самототожності. Мова є відображенням унікальності та самобутності нації, історичної тяглості змін у її становленні Водночас вона гостро реагує на суспільно-історичні події, а в окремих випадках стає предметом маніпуляцій задля різних політичних стратегій. Мова служить як засобом трансляції культурних кодів та цінностей, так і індикатором змін.

Вивчення актуалізації лексики української мови знаходиться в центрі уваги вчених-мовознавців. При цьому одне із значущих питань є вивчення змін в лексико-семантичній системі мови, зокрема, репресованої лексики. А саме, відродження забутих питомо українських слів, що використовувалися до 1930-х років і згодом були цензуровані. Як зазначає Ю. В. Шевельов: «Урядове втручання … у внутрішні закони мови було радянським винаходом і новиною. … радянська система встановлює контроль над структурою української мови: забороняє певні слова, синтаксичні конструкції, граматичні форми, правописні й орфоепічні правила, а натомість пропагує інші, ближчі до російських або й живцем перенесені з російської мови…[36, 3]». Мова йде про порушення мовної

картини світу українців, де насильницьке вилучення слів стало складовою її деформації.

Метою наукових розвідок мовознавців стало повернення автентичних українських слів, їхнього звучання та значення. Серед вітчизняних науковців, які розкривають зміст репресованої лексики, слід зазначити Ю. В. Антонів [2], О. М. Демську-Кульчицьку [6], С. Й. Караванського [8, 9], В. П. Кубайчука, Л. Т. Масенко [14, 15]. Так, О. М. Демською-Кульчицькою був створений Реєстр репресованих слів, в який ввійшло більше тисячі слів, що у 30-ті років XX століття були замінені або вилучені в результаті директивного «унормування» [6]. Ці лексеми детально аналізувалися також у дослідженні Д. В. Мазурик з метою розкриття внутрішнього механізму формування слів в українській мові [13]. Мовознавиці О. Г. Тулузаковою [34] та Л. Т. Масенко [16] дослідили репресовані слова, які використано у сучасних художніх творах та розмовній мові .

Отже, сьогодні з'явилася необхідність вивчення витіснених питомо українських слів, втрата/заміна яких викривила мовний розвиток українців, деформувала смислові значення, що зруйнувало мовну унікальність та самобутність. Адже, повернення в узус репресованої лексики сприяє відродженню культурної та національної самототожності. Однак незважаючи на зростання інтересу до репресованих слів, практичних онлайн-застосунків для їхнього вивчення та популяризації недостатньо.

Все це зумовило запит на створення онлайн застосунку, а саме чат-боту - програми, у якій здійснюється обмін повідомленнями між людиною та комп'ютером. Створення чат-бота у нашій роботі представлено інструментом актуалізації репресованої лексики для ознайомлення з ними. Оскільки надзвичайно популярним серед молоді є месенджер Telegram, то чат-бот було розроблено саме у цій

онлайн-платформі. Таке програмне забезпечення є доречним для залучення молоді до вивчення цього лексичного пласту. Ця можливість також особливо значима та цінна у такі важкі та відповідальні часи для українців як зараз - війну за право бути незалежними та суверенними.

Об'єктом дослідження є репресована лексика української мови та цифрові технології створення чат-бота.

Предметом дослідження є модель цифрового текстового контенту діалогової системи.

Мета роботи - створити чат-бот у месенджері Telegram, який виконує функцію ідентифікації репресованої лексики української мови.

Реалізація мети передбачає розв'язання низки *завдань:*

1) проаналізувати теоретичну мовознавчу літературу за темою лінгвоциду;

2) окреслити поняття «репресована лексика» та «реактивована лексика»;

3) провести лінгвосоціологічне опитування молоді для оцінки ставлення до репресованої лексики;

4) імпортортувати репресовані слова, їх відповідники та тлумачення до лінгвістичної бази даних за допомогою веб-скрапінгу;

5) створити код для автоматичної роботи чат-бота на платформі Telegram;

6) протестувати ефективність роботи чат-бота та внести корективи.

Для досягнення мети та виконання поставлених завдань, використано комплекс методів:

1) семантичний аналіз - метод визначення смислу та інтерпретації значення слів та фраз у контексті, в якому вони вживаються:

2) автоматизований контент-аналіз - метод кількісного вивчення великих обсягів текстів;

3) комп'ютерне моделювання - метод розв'язування задач із використанням комп'ютерних моделей;

4) веб-скрапінг - використовується для імпорту даних із веб-джерел;

5) метод аналізу даних - створення словника репресованих слів та їх сучасних відповідників через збір даних, їх очищення, аналіз та інтерпретацію результатів:

6) програмні бібліотеки: requests, beautifulsoup4, pandas, pymorphy2, python-telegram-bot. pymorphy2-dicts-uk.

Матеріал дослідження: Реєстр репресованих слів О. М. Демської-Кульчицької з 680 одиниць.

Практична та теоретична значущість результатів дослідження: отримані результати наукового дослідження виявили пробіл у популяризації репресованої лексики та її реактивації. Тому практичний результат полягав у створенні чат-боту на основі імпортованих із Реєстру репресованих слів О. М. Демської-Кульчицької [25]. Він створений у месенджері Telegram вперше та не має аналогів. Саме цим месенджером сьогодні активно користується молодь, що дає можливість зацікавити більшу аудиторію. Чат-бот спрямований на популяризацію репресованих слів та знайомство з ними, що є значущою практикою відновлення самобутності української мови. При цьому важливо повернути словам первинне смислове навантаження з урахуванням неповторних особливостей української літературної мови.

Структура роботи: кваліфікаційна робота складається з вступу, двох основних розділів та семи підрозділів, висновків, списку використаних джерел і додатків. Список використаних джерел нараховує 45 найменування, із них 8 - англомовні та 37 - україномовні, 15 - інтернет-джерела. Дослідження питомих рис української мови, вочевидь, є предметом пошуків українських мовознавців, що й зумовило малу представленість серед використаної літератури англомовних праць.

РОЗДІЛ 1

ДОСЛІДЖЕННЯ ЛІНГВОЦИДУ УКРАЇНСЬКОЇ МОВИ У СУЧАСНОМУ МОВОЗНАВСТВІ

1.1. Політика лінгвоциду української мови періоду 30-х років XX ст.

Сьогодні тривають активні розвідки з вивчення історичної пам'яті українців, що тісно пов'язано й з мовними трансформаціями. У центрі уваги життя та діяльність мовознавців того часу, які активно досліджували всі аспекти української мови. Багато хто з них став жертвами репресій. Інший напрям досліджень стосується коригування та цензурування низки словників української мови. Предметом дослідження також стали й слова, що вилучалися з активного вжитку та заборонялися. Вивчення репресованих слів, не лише відтворює події минулого, а й дозволяє зрозуміти та представити мовну спадщину. Крім того актуальним є ознайомлення з ними сучасників.

Сучасне мовознавство немає єдиного трактування для визначення слів, які використовувалися до 1930-х років та зазнали обмежень у використанні. Представлено такі поняття як «репресовані слова» та «репресована лексика» (О. М. Демська-Кульчицька [6]), «відроджені слова» (Ю. В. Шевельов [35]), «реактивовані слова» (Ю. В. Поздрань [20]), «внутрішні та зовнішні входження» (Д. В. Мазурик [13]), «заборонені слова» (Ю. В. Антонів [2]), «елімінована лексика» (Л. І. Мацько [17], М.А.Чаїнська [40]). Представлені поняття об'єднанні розумінням, що слова існували у мові впродовж тривалого часу і колись належали до активного словника, але згодом відійшли на задній план. Ці терміни є відображенням тенденції мови до перенесення слів від

периферійних позицій до центральних у межах лексичної системи. Важливим аспектом цього процесу є насильницьке вилучення слів з активного вжитку.

Отже, виявлено, що серед вітчизняних науковців немає чіткого розмежування між поняттями «репресовані слова», «відроджені слова», «елімінована лексика», «заборонені слова» й «реактивовані слова».

Із метою комплексного вивчення проблеми варто розглянути супутні явища, зокрема, лінгвоцид української мови. Окремі аспекти цього питання почали висвітлювали у своїх наукових розвідках ще І. І. Огієнко [18], М. І. Костомаров [11], І. Я. Франко [35].

Аналіз наукових досліджень свідчить про систематичну тривалу практику втручання у внутрішній розвиток української мови заради її штучного зближення з російською радянською владою. Документи свідчать, що цей процес був інтенсивним у 30-х роках та - у 50-60-х. Питомі риси української лексики, фразеології й навіть окремі граматичні форми поступово й послідовно замінялися кальками з російської мови. Таке явище науковці, вказуючи на насильницький характер змін у розвитку мови, назвали лінгвоцидом.

Першим, хто ввів у використання термін «linguicide», був український вчений Я. Б. Рудницький [26]. Так дослідник позначав спроби суспільства чи органів влади «обмежити використання однієї мови, натомість підтримуючи вживання іншої» [26, с. 68]. У своїй праці «Лінгвоцид» Я. Б. Рудницький детально описав цей процес, зазначивши що всі заходи, що ускладнюють природний розвиток мови, належать до явищ лінгвоциду [26].

Сьогодні явище лінгвоциду в Україні досліджується у контексті русифікації. Зокрема, С. Й. Караванський [8, 9] пов'язав поняття лінгвоциду з процесом зближення української мови та інших, зокрема, - російської.

Вивченням цієї проблеми також займалася А. П. Сваричевська. Вона показала, що історія пригноблення української мови триває впродовж більш ніж чотирьох століть, а головним чинником цього пригнічення є росія. А. П. Сваричевською окреслено основні форми лінгвоциду української мови з боку росії, зокрема, на основі лексичних змін: зниження статусу та авторитету мови; визнання мови неприродною; спроби «зближення» та уніфікації мови; обмеження соціального сприйняття мови; протидія «поширенню української мови» [27].

В. В. Лизанчук зібрала документальні матеріали лінгвоциду в Україні, розпочинаючи від середини XI століття [12]. А Л. Т. Масенко детально представила насильницьку мовну політику радянського керівництва в Україні [14, 15, 16].

На думку Я. К. Радевич-Винницького, «правда про нищення української мови має стати невід'ємною складовою історичної пам'яті українців, їхньої національної свідомості» [22, 3].

Я. К. Радевич-Винницький розглядав лінгвоцид як предмет українознавства, окреслюючи останнє як цілісну науку про Україну з етнологічними, психологічними, соціолінгвістичними, політичними, економічними, демографічними, регіональними, культурологічними, релігійно-конфесійними та освітніми аспектами [23].

У хрестоматії, упорядкованою В.В.Німчуком, зібрано праці науковців І. І. Сокола, Л. В. Ажнюк, Т. В. Майданович та інших, в яких представлено історію українського правопису впродовж XVI - XX століття. У ній розглянуто «традицію формування української правописної системи на наукових засадах із урахуванням неповторних особливостей української літературної мови» [7]. Прослідковано також тяглість мовних традицій та випадки штучно спровокованих змін.

На думку Н. І. Гавдиди та Л. Т. Назаревич, симіляційна політика радянської влади була значно жорсткішою, оскільки відбувалась на рівні

штучних змін у внутрішній структурі мови (переписували словники, фонетично уподібнюючи українські слова до російських), «а тому й більш результативною, ніж відповідна мовна політика колишньої царської Росії» [5, 79].

«Конфліктність теперішніх мовних проблем Україна успадкувала від колоніального минулого» - продовжують науковці [5, 79]. Так, російська мова витіснила українську в значній частині східних, південних та навіть деяких центральних областей України. Російська мова стала головним засобом спілкування. Такий мовний дисбаланс: «…послаблює національну самосвідомість українців, перешкоджаючи побудові міцної незалежної держави» [5, 79].

Радянська політика намагалась зупинити розвиток української нації у різних сферах, і придушення української лексики було однією з них. Зокрема, за період заборон було цензуровано цілу низку українських словників: «Російсько-український словник» О. П. Ізюмова; «Російсько-український словничок термінів природознавства і географії» К. В. Дубняка; «Російсько-український фразеологічний словник» В. П. Підмогильного та Є. П. Плужника (1927 рік); «Правописний словник» Г. К. Голоскевича (1929 рік); «Словник ділової мови» М. Ю. Дорошенка, М. О. Станиславського, В. М. Страшкевича; «Український стилістичний словник» І. І. Огієнка; «Російсько-український словник правничої мови» А. Л. Дроб'язка; «Російсько-український словник термінів фізики і хімії» М. П. Вікула; «Словник технічної термінології» І. М. Шелудька, Т. Садовського. О. Н. Синявський у «Нормах літературної української мови» (1941 рік) показував українцям як послуговуватися розмаїттям граматичних і стилістичних ресурсів рідної мови.

Словник «Російсько-український словник» (1924-1933 рр.) за редакцією А. Ю. Кримського та С. О. Єфремова, Ю. В. Шевельов назвав

«монументом українського культурного відродження 20-х років у царині мовознавства» [36]. Це перша академічна праця, яка у 20-30-х роках XX століття утверджувала норми української літературної мови та визначала її сферу використання.

У такий спосіб джерела мовного еталону ставали недоступні та табуйовані для використання.

Особливої уваги заслуговує Харківський правопис 1928-го року, відомий також як «правопис Скрипника». Таку назву він отримав на честь провідника радянської українізації М. О. Скрипника, який підтримував його впровадження та уніфікацію українського правопису. Харківський правопис встановив систему правил і норм для написання української мови у різних сферах життя в радянський період. Цей правопис був розроблений з участю провідних українських лінгвістів, у тому числі М. М. Грушевського, А. Ю. Кримського, О. Н. Синявського, О. Б. Курило, М. Г. Йогансена, С. О. Єфремова, М. Г. Хвильового, М. О. Ялового [32]. Він відрізнявся від попередніх правописів і спрямовувався на спрощення та уніфікацію правописної системи. У його рамках впроваджувалися нові правила щодо написання окремих літер та регулювання вживання деяких букв та орфографічних норм. Його вплив на українську мову та літературну культуру став помітним явищем того періоду, що не змогла не помітити радянська влада. Так Харківський правопис був підданий критиці за деякі несприйнятні чи обгрунтовані інновації та згодом був замінений більш консервативними та уніфікованими правилами.

Дослідження Харківського правопису 1928-го року засвідчує його значимущу роль у розвитку української мови в період радянської влади, розкриває етапи та трансформації українського правопису. Вже у 1933-му році його витіснив «правопис Постишева» - український правопис, максимально наближений до російської мови.

Лише у червні 2019 року в незалежній Україні було відновлено низку особливостей Харківського правопису. В. Ю. Васильєв наголошує, що таке нормування мало наукове підґрунтя - повернення до питомої української мови [4].

Українські мовознавиці, які брали участь у створенні Харківського правопису, згодом були репресовані. Самого М. О. Скрипника було доведено до самогубства. Була знищена ціла когорта мовознавців, тих, хто активно працював над розвитком української лексики. Ось деякі з імен. М. М. Сулима був відомий дослідженнями у галузі лексикографії та стилістики, О. Н. Синявський працював у галузі діалектології та історії української мови (цит. за [10]).

Отже, у сфері лінгвістики проблема лінгвоциду є важливим аспектом вивчення наслідків русифікації. Разом із цим її важливо розглядати комплексно, у контексті українознавства.

У працях О. Г. Тулузакової зустрічається питання «актуалізації слів», що трактується як збільшення активності вживання слів в мові. Цей процес пов'язаний з актуалізацією понять, які вони позначають, а також переміщенням слів із периферії мовної свідомості до активного словника [33]. Іншими словами, це повернення словесних одиниць із пасивного до активного.

Процес актуалізації словесних одиниць супроводжується змінами у їх значеннєвому обсязі, включаючи як розширення, так і звуження, переінтеграцію значень та трансформацію їхньої структури. Відродження слів, які раніше були забуті, та їх повторна активізація є помітним і значущим явищем у мові. Прикладами забутих або напівзабутих слів є: *світлина (фотографія), шпиталь (госпіталь), витоки (джерела), наклад (тираж)*. Ці слова здійснили переміщення з віддалених куточків мовного вжитку до активного словника. Більшість цих слів було виключено з

лексикону української мови штучним чином, навіть незважаючи на їхнє колишнє поширене вживання.

1.2. Репресована та реактивована лексика: трактування термінів

Поняття репресованої лексики використовує С. Й. Караванський. За його словами українську мову впродовж тривалого часу перетворювали на «бліду й незграбну копію російської мови» [8, 5]. Таке уподібнення відбулося внаслідок «злиття мов», з метою знищення її самобутніх ознак.

Терміни «репресовані слова» та «репресована лексика» досліджувала українська лінгвістка О. В. Демська-Кульчицька [6]. Вона пояснює ці терміни як слова, або ж вирази, що були цензуровані в українській літературі та мовленні у певний історичний період у зв'язку з політикою та ідеологією радянської влади. Такі слова вважалися неприйнятними для використання у формальному мовленні та офіційних словниках. Вони замінювались на російські відповідники або взагалі заборонялись. Розглядаючи терміни «репресоване слово» та «репресована лексика» О. В. Демська-Кульчицька також вказує на їх змінність, адже список заборонених слів формувався залежно від політичних, ідеологічних та культурних чинників у конкретний історичний період [6]. Як наслідок, у різний історичний період слова та вирази, які були об'єктом цензури відрізнялись. У своїх дослідженнях лінгвістка по-різному трактувала ці терміни, залежно від їхньої складності. Поняття «репресоване слово» включає в себе окреме слово, яке було виключене з публічного вжитку чи цензуроване у літературі. Натомість «репресована лексика» - це не лише окремі слова, але й фрази, вирази, терміни та інші лінгвістичні одиниці, які піддавалися цензурі або забороні [6].

Мовознавець Ю. В. Шевельов досліджував фонологію української мови. Він розглядав відродження цензурованих слів як прояв

національної свідомості та боротьби за національну самостійність у різні епохи української історії. Науковець зазначав: «Моє давнє внутрішнє неприйняття радянщини примхливою психологічною грою ввійшло в асоціяції зі станом української культури: її женуть і переслідують, значить - ми спільники» [36, 32]. Діяльність Ю. В. Шевельова є ілюстрація того, як примус і терор викликали протилежну реакцію захисту та дії.

У своїх працях Ю. В. Шевельов трактував термін «відроджені слова», як слова, що відновлюються в мовленні та літературі після періоду політичних репресій [36]. Мовознавець віддавав увагу не лише самим словам, а й їхньому значенню і символіці у контексті української історії та культури. Він досліджував, як «відроджені слова» відображають боротьбу за мовну свободу та національну ідентичність.

Українська мовознавиця Ю. В. Поздрань розглядала поняття «реактивовані» у контексті відновлення слів чи виразів, які раніше втратили своє активне вживання в мовленні. У своїх дослідженнях Ю. В. Поздрань вивчала забуті або ж маловживані слова, які відроджуються з метою становлення національної ідентичності українців. Термін «реактивовані» вона трактувала як відновлені у мовленнєвій практиці та літературі слова [20]. Вона зазаначала, що «Процес реактивації лексики сприяє розширенню словника української літературної мови власними надбаннями й засвідчує зростання інтересу до національної мови» [20, 48].

Л. І. Мацько оперує терміном «елімінована лексика», тобто вилучена з активного вживання з певних причин [17].

Д. В. Мазурик, українська мовознавиця, використовувала термін «внутрішні та зовнішні входження» щодо способів, якими нові слова або лексичні одиниці потрапляють до мови. «Внутрішнє входження» - це процес утворення нових слів або лексичних одиниць в межах самої мови за допомогою мовних засобів, які вже існують у цій мові [13]. Тобто

слова, які стали об'єктом політичних або культурних репресій, можуть відновлюватися в мовленні через процеси внутрішнього словотворення. Процеси внутрішнього словотворення включають афіксацію, що полягає у формуванні слів за допомогою додавання префіксів та суфіксів; складання, що передбачає об'єднання двох або більше слів для утворення нового; абревіацію, яка полягає в скороченні слова шляхом відкидання складових; словоскладання, утворення нових слів з існуючих без додавання афіксів; конверсію, яка передбачає перехід слова з однієї частини мови в іншу без зміни форми; неосемантизація, що зумовлює зміну значення слова внаслідок внутрішнього семантичного переходу; внутрішнє синтаксичне словотворення, яке включає утворення нових слів шляхом зміни порядку слів або використання специфічних синтаксичних конструкцій; інші процеси, такі як редуплікація та зворотне словотворення.

Ці процеси детально аналізуються у дослідженні Д. В. Мазурик з метою розкриття внутрішнього механізму формування слів в українській мові [13]. Тоді як поняття «зовнішнє входження» характеризує процес, коли нові слова або лексичні одиниці потрапляють до мови з інших мов або джерел. Процеси зовнішнього словотворення включають запозичення, тобто введення слів з інших мов, адаптацію, яка передбачає пристосування запозичених слів до власної фонетичної, морфологічної та семантичної системи мови, калькування, що полягає у використанні власних слів за зразком іноземних, та інші процеси, такі як калькування, рефлексія та інші. Наприклад, під час періодів політичного тиску можуть бути запозичені нові терміни або фразеологізми з інших мов як засіб уникнення цензури [13].

Окремий аспект представлено спробую заміни одних слів іншими з метою певного дистанціювання від російської мови. О. І. Бондар процес

повернення до активного вжитку лексики, забороненої в тридцяті роки, називає «реабілітацією словникового складу української мови» [3].

Загалом, ми вважаємо, що доцільно трактувати поняття «репресована лексика» як слів, які були насильницько вилученні/замінені з активного вжитку українців радянською владою. А реактивовані слова трактуються як відновлені у мовленнєвій практиці та літературі репресовані слова, які відроджуються з метою національної самототожності українського етносу

Хід дослідження вимагає уваги до сфери вживання репресованих слів та їхнього тлумачення. Варто прослідкувати спроби їхньої класифікації для детального аналізу.

1.3. Реактивація репресованої лексики у процесі формування національної самототожності українського етносу

Вивчення проблеми «репресованих слів» прослідковується у працях В. П. Агєєвої [1], Ю. В. Антонів [2], О. М. Демської-Кульчицької [6], С. Й. Караванського [8, 9], Л. Т. Масенко [14, 15], Л. І. Мацько [17], М. А. Чаїнською [40]. Метою наукових розвідок було повернути автентичність українських слів, їхнє звучання та значення.

Встановлення реєстру «репресованих слів» провела Ю. В. Антонів, проаналізувавши редакторські правки робочого примірника перекладу роману В. Г. Еша «Вибір зброї» (1966 рік), зробленого О. Д. Сенюк. За результатами дослідження встановлено, що найактивніше з української мови витіснялася лексика західноукраїнського походження [2]. Українські лексеми називалися штучними, архаїчними та провінційними. Окремі лексеми були визнані націоналістичними.

Особлива увага при цьому надавалася термінам у галузях науки, промисловості й сільського господарства. Так, заміна відбулася у технічних термінах: *«автомобілярню»* почали називати автозаводом; *«далекогляд»* - телескопом; *«мутра»* - гайкою; *«електровня»* -

електростанцією. Спеціальними комісіями було переглянуто математичну, фізичну, географічну й ботанічну термінологію та підготовлено п’ять бюлетенів для замін. Маятник українці називали *«вагалом»* або *«хитуном»*, діагональ - *«перекрутнею»*, перпендикуляр – *«простопадом»*, континент - *«суходолом»*, западина - *«улоговина»*. Полюс називали *«бігуном»*, а полярне сяйво - *«північною загравою»*, вічна мерзлота була «мерзлиною», затемнення – *«міненням»,* а туманність – *«мряковиною»*. Коли щось брали в дужки, говорили *«заклямовувати»*.

Як зазначає О. М. Демська-Кульчицька: «Розвиток української мови впродовж усього ХХ століття мав частіше насильницький, аніж еволюційний характер» [6, 34]. Свою позицію мовознавиця підтверджує створенням Реєстру репресованих слів. До нього було включено перелік лексичних одиниць української мови, які у різний час були замінені або вилучені в результаті директивного «унормування». Дослідивши архіви, вчені визначили викинуті слова з текстів у 1920-ті роки, потім у 1960-ті роки та 1980-ті роки. Основою словника-реєстру стали загальновживані та термінологічні слова, які зазнали повної чи часткової заміни. Слова могли змінюватися й повністю, наприклад, *двосічна* стає бісектрисою, *підступ* замінюється на послідовність, *рівник* на екватор, *відтак* на потім, занапастити на погубити, гостриця на піраміда, надма на дюна, або цілком вилучалися, як бігме, либонь, лицедій. Також змін зазнали префіксально-суфіксальні утворення, що не збігалися з «мовними правилами», наприклад, заміна *«прихистити»* на «захистити», а також через «неправильне» синтаксичне використання слів, наприклад, *«розсіювання світла»*, яке стає «світлорозсіюванням». Так відбуваються лексико-семантичні деформації слів.

У Реєстрі фігурують слова, для яких не існують жодні альтернативи. Відсутність замінників розглядається як ще один метод

зменшення різноманітності слів у мові. Такий підхід передбачав не просто введення нових термінів, які характерні для інших мов, а також видалення вже існуючих лексем з мовного арсеналу. Наприклад, слова *«завше», «всенький», «достоту», «походеньки», «навдивовижу»* стали жертвами цього підходу.

Зазначимо, що частина цих слів зникли б з активного вжитку самостійно як застарілі, а частина слів - відновились попри цензуру. До активного вжитку питомих українських слів повернулися: *шпиталь, благодійник, злука, світлина, вар'ят, часопис, летовище.*

Реєстр репресованих слів є веб застосунком, призначеним для зберігання, обробки та надання доступу до інформації про слова, що були заборонені або витіснені з української мови в різні історичні періоди. Він має загальну структуру, що включає три розділи: перше поле містить вихідне слово разом з коментарем, що пояснює причину його усунення або заміни; друге поле містить слово-замінник та пояснення мотивації заміни; третє поле містить посилання на текстові джерела, що рекомендують або зафіксовують зміни в лексиконі. При цьому встановлено, що коментарі надзвичайно однотипні та поверхові.

У рубриці «Покликання» скорочено подано прізвища автора та рік виходу праці з поясненням необхідності заміни чи вилучення певних слів. Найчастіше згадано партійного діяча А. А. Хвилю, першого автора правок українського правопису та наближення української мови до російської.

Реєстр став предметом активного дослідження українських мовознавців. Існують різні класифікації репресованих слів. Реєстр насамперед включає терміни і загальновживану лексику. О. Г. Тулузакова представила класифікацію іменників загальновживаної лексики, що стосуються щоденних реалій життя пересічної людини [34]. Йдеться про поняття, пов'язані з житлом, його організацією, а також характеристики

окремої особи, включаючи психічні та фізичні властивості. Класифікація відбувалась на основі введення репресованих слів у Словник української мови в 11 томах [29]. Перша група охоплює слова, які зафіксовані у Словнику без стилістичного обмеження. Друга група - лексеми, що мають певні обмежувальні ремарки. Оскільки словник був створений у період з 30-х років до 70-х років XX століття, це дало можливість відслідкувати вилучення лексики з цього періоду, а також зафіксувати обмеження у вживанні слів.

О. Г. Тулузакова представила приклади актуалізації репресованої лексики у прозових творах Ю. І. Андруховича, Ю. П. Винничука, Ю. Р. Іздрика. Так, *виправа* означає *подороож, мандрівку, похід; громада - організацію; гурт - група; гурток - група; обрис - контур; загал - сукупність; крамниця - магазин; припис - правило; світлина - фотографія; слоїк - банка; слухавка - телефонна трубка; хідник - тротуар; цигарки - сигарети; шпиталь - госпіталь* [34]. Як бачимо, навіть у директивний спосіб не вдалося повністю усунути питомо українську лексику.

Актуалізувалися лексеми, що, по-перше, отримали ремарку фамільярного та розмовного: *білявка* - світловолоса; *дзиґлик* - стільчик; *зух* - молодець; *краля* - красуня; *курдупель* - коротун; *стрій* - наряд; по-друге, що зазначалися як застарілі: *дзиґарі* - годинник; *добродій* - дорослий чоловік; *мапа* - карта; *шинквас* - бар; по-третє, що отримали маркування діалектне: *люстро* - дзеркало; *обрус* - скатертина; *робітня* - майстерня; по-четверте, з маркуванням рідко: *однострій* - уніформа; *таця* - піднос; по-п'яте, як аналог: *вивірка* - білка; *карафка* - графин; *крамар* - торговець; *філіжанка* - чашка; *холодник* - холодильник; *часопис* - газета; *церата* - клейонка; *вар'ят* - божевільний; *пігулки* - таблетка. Так, частина штучно вилучених слів активізувалася у художніх творах через внутрішньомовні ресурси.

Л. І. Мацько зазначає, що повернення до активного словника української мови силоміць вилучених елементів з норми допоможе «відтворити і поповнити національний образ нашої мови на всіх її структурних рівнях завдяки приведенню у відповідність з етнічною природою української мови її літературних норм» [17, 15].

У такий спосіб, представлено явище реактивації репресованих слів в сучасній українській усній та писемній мові. Представлено приклади актуалізації репресованої лексики у прозових творах.

1.4. Суспільна оцінка лінгвоциду української мови: соціолінгвістичний експеримент

Ставлення до мови є одним із чинників творення нації та проявом ідентичності. Також воно вказує на загальні оцінки, які існують у суспільстві щодо мови: престижність, традиції, розповсюдженість, статусність. Українське мовне середовище має власні специфічні характеристики, а вивчення ставлення до мови постає маркером змін в суспільстві. Знання про історію деформації української мови, зокрема, про репресовану лексику, є грунтом для раціонального сприйняття мови як значущого чинника національної ідентифікації.

В опитуванні взяло участь 68 студентів рівня фахової вищої освіти закладу вищої освіти міста Києва, віком 17-18 років. Воно проводилося у квітні 2024 року. До анкети, на основі якої проводилося опитування, увійшло 11 запитань. Опитувальник було створено з використанням додатку Google-форма та проводилося за принципом анонімності (Додаток 1).

Проведене опитування «Ставлення молоді до репресивної мовної політики радянської влади» вивчало два аспекти: по-перше, обізнаність

молоді з питанням репресованої лексики, по-друге, усвідомлення негативних наслідків мовного лінгвоциду [8].

За результатами опитування українська мова є домінуючою мовою їхнього повсякденного спілкування. Так, 62% опитаних спілкуються українською мовою, 9% - російською мовою, 29% - українською та російською (рис.1.1). Українська мови усвідомлюється рідною - є чітким маркером самоусвідомлення власної національної, а разом із тим і мовної ідентичності.

Рис. 1.1. Діаграма відсоткового розподілу відповідей на запитання «Якою мовою ви спілкуєтесь у повсякденному житті?»

Рис. 1.2. Діаграма оцінок рівня володіння українською мовою

При цьому 85% респондентів у своїх відповідях зазначають свій рівень володіння українською мовою як високий, оцінивши його таким чином: 10 балів - 16,2%, 9 балів - 30,9, та 8 балів - 38,2 (рис. 1.2). Такий показник свідчить про високу самооцінку мовної компетентності, відсутність страху перед мовною невдачею, комплексів щодо власного мовлення.

Ілюстрацією обізнаності респондентів із проявами лінгвоциду, які проводила радянська влада задля переходу українців на російську мову, є знання про репресовані слова. Згідно опитування 91% респондентів відома така ситуація, вони відповіли ствердно, що знають про факт репресивної мовної політики (рис. 1.3).

Однак, назвати такі слова може лише 10% респондентів, частково їх може назвати - 64%. При цьому 26% респондентів таких слів не знають (рис. 1.4).

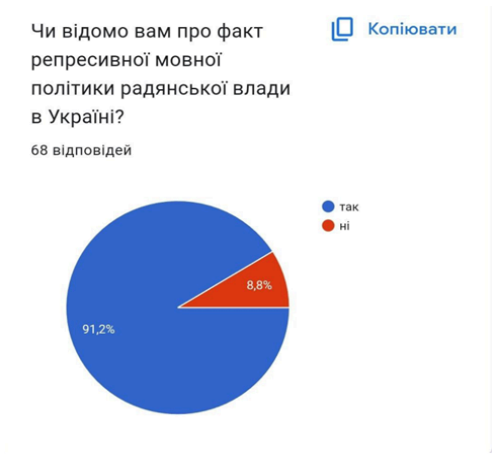


Рис. 1.3. Діаграма оцінок рівня знань репресованої лексики

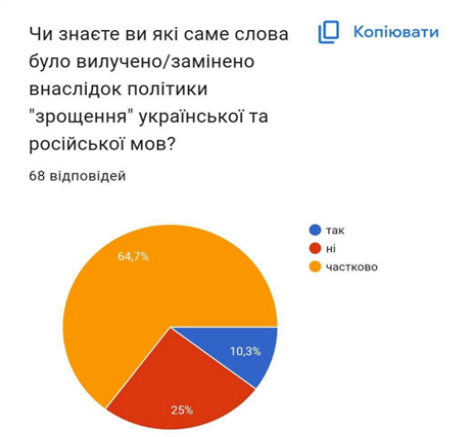


Рис. 1.4. Діаграма розподілу відповідей на запитання «Чи знаєте ви які саме слова було вилучено/замінено внаслідок політики «зрощення мов».

Цікаво прослідкувати чи «впізнають» респонденти репресовані слова. Слово «підсоння» не знають 87% респондентів, а слово «оник» - 47% (рис. 1.5, 1.6). Як бачимо, співзвучні слова у більшості випадків незрозумілі респондентам. Тобто, частина слів та значень є витісненими з мовної свідомості українців.

Рис. 1. 5 Діаграма розподілу відповідей на запитання «На вашу думку, що означає слово «підсонння»?

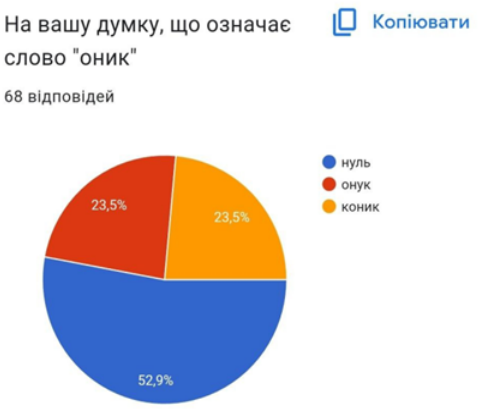


Рис. 1. 6 Діаграма розподілу відповідей на запитання «На вашу думку, що означає слово «оник»?

Виявлено високий рівень переконаності серед молоді, що в Україні існує мовна проблема. Так, 87% респондентів у своїх відповідях вказали, що мовна проблема існує (рис. 1.7).

Наслідками репресивної мовної політики радянської влади респонденти вважають (у порядку спадання): збіднення української мови (72%), штучне зближення української та російської мов (66%), розвиток двомовності (44%), збагачення української мови (10%), задоволення потреби у нових термінах індустріального суспільства (6%). П’ятеро

респондентів вважають таку ситуацію надуманою проблемою (рис. 1.8).

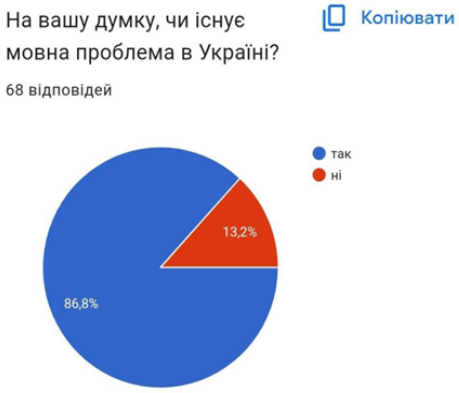


Рис. 1.7 Діаграма розподілу на запитання «На вашу думку, чи існує мовна проблема в Україні?»

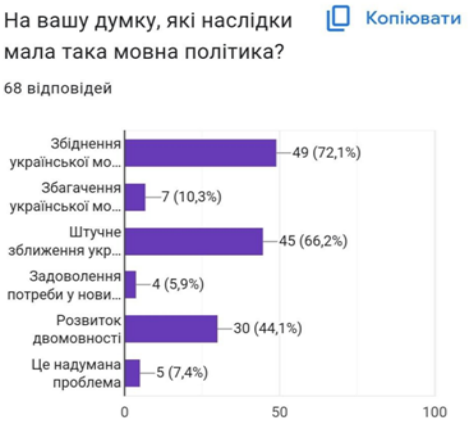


Рис. 1.8 Діаграма розподілу відповідей щодо наслідків мовної політики радянської влади

При цьому, вплив репресивної мовної політики радянської влади на мовну ситуацію оцінюється на середньому рівні у 32% відповідях та на високому - у 63% (рис. 1.9). Так, у мовній картині світу сучасної молоді представлено негативний аспект цензурування української мови у радянському минулому.

Позитивно респонденти ставляться до можливості вживання репресованих слів у медіа 38% респондентів, тоді як 27% – негативно. Для 35% респондентів це питання нецікаве та неактуальне (рис. 1.10).

Рис. 1.9 Діаграма оцінок впливу репресивної мовної політики радянської влади на мовну політику

Рис. 1.10 Діаграма розподілу відповідей на запитання «Як би ви зреагували на вживання репресованих слів у медіа?»

Серед респондентів домінує думка, що українська мова та російська не є найбільш спорідненими (рис. 1.11). Ствердну відповідь дало 94% респондентів. Лише 6% опитаних вважають інакше. Так, один із російських аргументів, що українці та росіяни – один народ і тому мають жити в одній державі, не має вагомої переваги та підтримки. Однак все ж присутній в мовному дискурсі молоді.

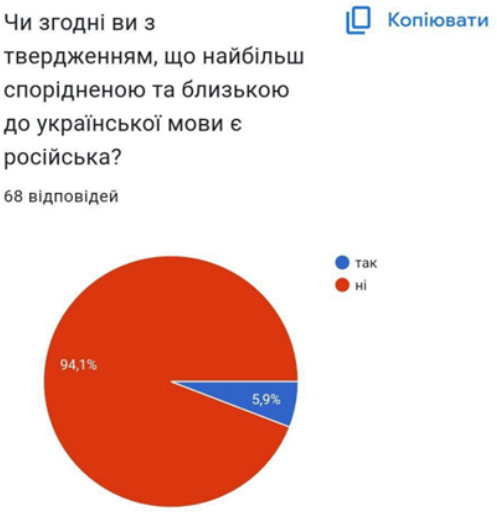


Рис. 1. 11 Діаграма оцінок спорідненості української та російської мов

За результатами проведеного опитування «Ставлення молоді до репресивної мовної політики радянської влади» встановлено, що здобувачам вищої освіти відома ситуація нищення української мови радянською владою та існування репресованих слів [8]. Це стало невід'ємним складником історичної пам'яті українців, їхньої національної та мовної самосвідомості. Також яскраво виражене негативне ставлення

до цензурування української мови у радянському минулому. Отже, вивчення репресованої лексики є актуальною темою для сучасної молоді.

Зазначимо, що згідно з опитуванням, українська мова є домінуючою мовою повсякденного спілкування молоді. Така ситуація засвідчує високий рівень мовної національної свідомості, функціонально-побутового характеру української мови для більшості молоді.

Висновки до першого розділу

Таким чином, низка вітчизняних мовознавців присвятили свої праці дослідженню проблеми цензурування української мови й заклали теоретичне підґрунтя для подальшого вивчення мовного дискурсу. У дослідженнях представлено негативні явища у розвитку української мови як заміна/вилучення питомо українських слів. Явище свідомого нищення української мови отримав усталену назву - лінгвоцит. Внаслідок цензурування радянською владою української лексики зникло безліч українських лексичних одиниць, які було витіснено на периферію активного словника, а також чимало питомих мовних рис. Сьогодні явище лінгвоциду в Україні представлено у контексті русифікації.

Лінгвоцид розглядається комплексно, у контексті українознавства. Тоді як термін «репресована лексика» є предметом дослідження лінгвістики та її наукового підходу і практичного застосування. Вважаємо, що доцільно трактувати поняття «репресована лексика» як слів, які були насильницько вилученні/замінені з активного вжитку українців радянською владою. Це відбувалося шляхом фонетичного уподібнення українських слів до російських. Таке формулювання точно передає їхню характерну ознаку - насильницьке вилучення українських слів з активного вжитку, цензерування з метою знищення самобутності

української мови, фонетичного уподібнення українських слів до російських.

Разом із тим спостерігається реактивація репресованих слів у художніх творах та розмовній мові. Встановлено, що такі зміни зумовили суттєві лексико-семантичні зсуви в смислових інтерпретаціях українських слів.

Проаналізовано Реєстр репресованих слів, створений О. М. Демською-Кульчицькою. До нього ввійшов перелік лексичних одиниць української мови, які у різний час були замінені або вилучені в результаті директивного «унормування», та віднайдено їхнє лексичне значення (близко 1000 слів). Також проаналізовано сферу вживання репресованих слів: загальновживані та термінологічні слова. Даний реєстр став основою для створення чат-бота. Всього було імпортовано 680 слів.

У ході дослдіження також представлено результати опитування про ставлення молоді до репресованих слів та усвідомлення негативних наслідків мовного лінгвоциту. Констатовано яскраво виражене негативне ставлення молоді до цензурування української мови у радянському минулому та зростання інтересу до звучання та значення слів, заборонених/витіснених радянською ідеологією.

Витіснення та цензурування питомо українських слів трактується молоддю як свідома мовна політика радянської влади - знищення значущої ознаки нації. Наслідком вважається збіднення української мови, штучне зближення української та російської мов, розвиток двомовності.

Повернення та реактивація репресованих слів є значущою практикою відновлення самобутності української мови. При цьому важливо повернути їм первинне смислове навантаження з урахуванням неповторних особливостей української літературної мови. Інструментом

актуалізації репресованої лексики є розроблення чат-бота в месенджері Telegram, про що мова піде у наступному розділі.

РОЗДІЛ 2

СТВОРЕННЯ ЧАТ-БОТА “ПОВЕРТАЄМО РЕПРЕСОВАНУ МОВУ”

Дослідження теоретичних аспектів репресованої лексики тісно пов’язано з застосуванням їх на практиці, в якій особливе місця займає вивчення репресованої лексики, ознайомлення із проблемою відновлення втраченого лексичного пласту української мови у віртуальному інформаційному просторі. На наше переконання, найефективніший спосіб виконання цих завдань - це розроблення інтерактивної системи у формі чат-бота. З ним можна ознайомитись за покликанням: https://t.me/repressed_words_bot .

2.1. Технології створення та класифікації діалогових систем

Діалогові системи, як напрям штучного інтелекту, використовуються для взаємодії з користувачем засобами мови. Вони володіють можливістю розпізнавання та розуміння мовних команд або запитів, що надходять від користувача, та відповідають на них відповідно до своїх програмних алгоритмів.

Діалогові системи можуть використовувати різноманітні методи для забезпечення ефективної взаємодії з користувачем, такі як: правила, машинне навчання або нейронні мережі. Вони можуть впроваджуватися у веб-сервісах, месенджерах, програмах для особистих асистентів та інших платформах зв'язку, щоб надавати користувачам різноманітні послуги, а також допомагати їм у вирішенні певних завдань [38].

Класифікація діалогових систем може проводитися за різними критеріями. За функціональністю вони поділяються на інформаційно-пошукові системи, які надають інформацію за запитом (наприклад, пошукові системи), системи підтримки клієнтів, які

допомагають вирішувати проблеми клієнтів (наприклад, чат-боти для технічної підтримки), системи виконання завдань, що допомагають користувачам виконувати певні завдання (наприклад, голосові помічники), та системи навчання, що використовуються для навчання та тренування (наприклад, освітні платформи) [37].

За архітектурою діалогові системи поділяються на rule-based системи, які використовують заздалегідь визначені правила для обробки запитів, та data-driven системи, які використовують машинне навчання для обробки та генерації відповідей.

За способом взаємодії діалогові системи поділяються на текстові, які взаємодіють з користувачем через текстові повідомлення, та голосові, що використовують голосові команди для взаємодії.

Форма чат-боту - це одна з конкретних реалізацій діалогових систем, яка використовується для взаємодії з користувачами у вигляді чату або текстового спілкування [21].

Чат-боти можуть мати різні рівні складності та функціональності, від простих ботів, які відповідають на певні запитання або команди, до складних систем з розвинутим штучним інтелектом [28].

Ми обрали створення простого боту, оскільки такий підхід є оптимальним для поставлених цілей. Наша мета полягає в забезпеченні користувачам зручного та швидкого доступу до інформації про репресовану лексику. Створення боту, який може відповідати на конкретні запитання про репресовані слова та їх відповідники за допомогою готових фраз, дозволить користувачам легко знайти необхідну інформацію. Такий підхід сприятиме пізнанню та вивченню репресованої лексики, що є основним завданням роботи.

Отже, для забезпечення актуалізації репресованої лексики варто використати практичний інструмент, а саме, розроблення чат-бота в месенджері Telegram. Він відрізняється від теоретичного представлення

тим, що надає можливість користувачам дізнаватися про репресовані слова через онлайн-застосунок.

Перший чат-бот було створено у 1966 році програмістом Дж. Вейзенбаумом. Він розробив програму під назвою «ELIZA», яка була заснована на простій моделі обміну повідомленнями, де кожна фраза користувача аналізувалася та генерувалася відповідь [21].

Чат-боти можна класифікувати за кількома параметрами, такими як галузь знань, надання послуг, мета, метод обробки даних і генерування відповідей [37].

За класифікацією на основі області знань, чат-боти поділяються на боти відкритого домену (Open Domain), які можуть говорити на загальні теми, і боти закритого домену (Closed Domain), що зосереджені на певній галузі нань. За класифікацією на основі послуг, чат-боти поділяються на міжособистісні (Interpersonal), які надають комунікаційні послуги, як-от бронювання ресторанів або авіарейсів, і внутрішньо особистісні (Intrapersonal), які діють як особисті супутники користувача в додатках на кшталт Viber, Telegram та WhatsApp. Внутрішньо особистісні боти запам’ятовують інформацію про користувача. Міжагентські (Inter-agent) чат-боти забезпечують комунікацію між собою, наприклад, через інтеграцію Alexa-Cortana [43].

Класифікація за цілями враховує основну мету чат-ботів. Інформаційні (Informative) чат-боти надають користувачам інформацію, яка заздалегідь зберігається або доступна з визначених джерел, наприклад, чат-боти для відповідей на поширені запитання. Розмовні чат-боти (Chat-based/Conversational) спрямовані на спілкування з користувачем як людина, відповідаючи на отримані повідомлення [44].

Чат-боти, орієнтовані на виконання завдань (Task-based), призначені для виконання конкретних завдань, таких як бронювання авіарейсів чи ресторанів, або надання допомоги з пошуком інформації.

Класифікація за методом обробки вхідних даних і генерації відповідей враховує підходи до обробки даних і створення відповідей. Використовуються три основні моделі: модель на основі правил, модель на основі пошуку та генеративна модель.

Модель на основі правил (Rule-based) обирає відповіді на основі фіксованих попередньо визначених правил, заснованих на аналізі лексичних форм введеного тексту. Ці правила закодовані вручну та організовані в розмовні моделі, що дозволяє чат-боту відповідати на більше різновидів введених даних, але вони чутливі до орфографічних і граматичних помилок.

Модель на основі пошуку (Retrieval-based) забезпечує більшу гнучкість, оскільки вона шукає і аналізує доступні ресурси через API, витягуючи потенційні відповіді з індексу та використовуючи підхід збігу для вибору відповідної відповіді. Генеративна модель (Generative) створює відповіді на основі поточних і попередніх повідомлень користувача за допомогою алгоритмів машинного навчання та методів глибокого навчання, що забезпечує більш природні відповіді, хоча їхня розробка та навчання є складнішими.

Класифікація за участю людини визначає чат-боти за рівнем участі людини. Human-aided чат-боти використовують втручання людини в деяких компонентах, наприклад, через працівників, які покращують логіку роботи чат-бота.

Хоча такі чат-боти є більш гнучкими і надійними порівняно з повністю автоматичними системами, вони обробляють інформацію повільніше, що ускладнює масштабування для великої кількості користувачів.

Класифікація за відкритістю вихідного коду поділяє платформи на відкриті, як RASA, і приватні, як ті, що пропонуються Google або IBM. Відкриті платформи дозволяють розробникам втручатися в більшість

аспектів реалізації, тоді як закриті діють як чорні ящики, обмежуючи доступ до деяких функцій. Проте закриті платформи можуть мати перевагу завдяки доступу до великих обсягів даних, які збирають великі компанії. Разом з тим, є чат-боти, що поєднують кілька класифікацій одночасно [44].

Таким чином, чат-бот для надання інформації про репресовану лексику, є прикладом бота закритого домену через обмежену галузь знань, яку він охоплює.

Чат-боти можна розділити на два типи: скриптові і засновані на штучному інтелекті (AI). Скриптові боти використовують заздалегідь визначені сценарії з бібліотеки відповідей, тоді як AI-боти користуються NLP. Вони навчаються на запитах користувачів і, з причин вищої ефективності, набирають більшої популярності [43].

Існують також боти на основі кнопок, що ідеально підходять для ситуацій із чітко визначеними сценаріями, такими як відділ продажів чи підтримки. Ці боти пропонують вибір кнопок і запитують дані, а їх створення є легким і доступним.

Гібридний бот поєднує функціонал кнопкового та AI-ботів, дозволяючи вибрати кнопки або ввести питання. Він може бути корисним для розв'язання поширених питань, але може вимагати додаткової розробки.

Голосові асистенти, засновані на розмові, є найпрогресивнішим типом ботів, що потребують високої експертизи та команди спеціалістів. Вони прогнозуються як найпопулярніший вид у майбутньому. Незалежно від типу, чат-боти допомагають організаціям пришвидшити процеси, зменшити витрати та покращити сервіс.

Із метою створення бота, який ефективно взаємодіятиме з користувачами на тему репресованих слів, варто обрати використання скриптового типу бота. Цей вибір обумовлений потребою обмежити

діапазон можливих відповідей до конкретної теми. Скриптовий бот дозволить попередньо визначити набір правил і шаблонів відповідей, що забезпечить швидку та точну реакцію на запитання користувачів.

Крім того, обраний метод обробки даних за правилами дозволить попередньо визначити сценарії взаємодії з користувачем і створити набір відповідей на основі цих правил. Це забезпечить швидкий відгук на запитання користувачів і збереже точність та якість відповідей.

Ще однією перевагою скриптового бота є його здатність забезпечити конфіденційність та безпеку. Обмежений діапазон можливих відповідей допоможе уникнути ризику витоку конфіденційної інформації чи непередбачених ситуацій у взаємодії з користувачами [42].

Нарешті, використання скриптового бота не виключає можливості майбутнього розширення його функціональності. Поступово можна буде додавати нові правила та шаблони відповідей, щоб розширити можливості бота і покращити його взаємодію з користувачами у майбутньому.

Загалом, кожен тип чат-бота має свої особливості та переваги, що визначають його використання та ефективність у певній сфері діяльності.

Таким чином, можемо зробити висновок, що використання чат-бота є доцільним для надання інформації про репресовану лексику. При цьому доцільно використати бот закритого домену та скриптовий тип.

2.2. Проєктування чат-бота та його функціонал

Розробка чат-бота включає ряд етапів: від обробки запитів користувача до вибору платформи для реалізації [28]. Проєктування починається з визначення функцій бота, вибору методів обробки інформації та визначення його дизайну. Розуміння категорії чат-бота допомагає розробникам обирати правильні алгоритми. Основні етапи

включають розуміння запиту, виконання дій та генерацію відповіді. Вибір платформи для розробки може бути важливим кроком існують багато варіантів як комерційних, так і відкритих.

Сьогодні чат-боти підтримують більшість популярних месенджерів, один із них - Telegram. Це безкоштовний багатоплатформний додаток, розроблений на мовах програмування C++ та Java. Месенджер дозволяє користувачам обмінюватися повідомленнями і здійснювати дзвінки, надсилати файли різних форматів, створювати ботив, вести власні канали тощо. Клієнтські програми Telegram доступні для Android, iOS, Windows Phone, Windows, macOS та GNU/Linux. Станом на серпень 2023 року кількість щомісячних активних користувачів сервісу перевищила 800 мільйонів осіб [45]. Окрім обміну повідомленнями в чатах і групах, месенджер дозволяє зберігати необмежену кількість файлів, вести канали (мікроблоги), а також створювати і використовувати ботив.

Для створення чат-бота, який здатен розуміти та взаємодіяти з користувачами українською мовою, необхідно зібрати великий обсяг лексичних даних у базу даних. Ручне збирання цих даних є надзвичайно трудомістким та часозатратним процесом. Веб-скрапінг, тобто автоматизоване збирання даних із веб-сторінок, є ефективним рішенням для цього завдання. Веб-скрапінг дозволяє автоматично витягувати інформацію з веб-сторінок і зберігати її в структурованому форматі, такому як таблиця Excel або CSV. Це значно пришвидшує процес збору даних та мінімізує ймовірність помилок, пов'язаних із ручним введенням [42].

Було розроблено технічне завдання, яке передбачало таку послідовність дій:

1. Написання коду [1], для автоматичного імпорту репресованих слів та їх відповідників із реєстру репресованих слів, створеного

О. В. Демською-Кульчицькою [25]. Очищення отрманих даних проходило вручну.

2. Автоматичний імпорт тлумачень [2] з СУМ у 11 томах [28]. Пошук тлумачень у СУМ у 20 томах [29] та редагування отриманої інформації вручну.
3. Створення чат-бота [3]:
    - Завантаження списку репресованих слів та їх сучасних відповідників із створеного CSV-файлу.
    - Встановлення бібліотек «pandas» [1], «pymorphy2» [4] і «python-telegram-bot» [7].
    - Використання морфологічного аналізатора «pymorphy2» [4] для отримання всіх можливих форм репресованих слів та їх сучасних відповідників.
    - Написання функцій для обробки текстових повідомлень від користувачів, розбивання їх на окремі слова та пошук репресованих слів серед них.
    - Додавання обробника команди «/start» для привітання.
    - Запуск бота у режимі очікування повідомлень від користувачів та обробки їх запитів.
    - Тестування бота з різними варіантами введених повідомлень та виправлення помилок.

Всі коди створено у середовищі Google Colab [9].

На початку роботи було створено код, що допомагає імпортувати репресовані слова та їх відповідники автоматично [1]. Дані були взяті із рєстру репресованих слів, створеного О. В. Демською-Кульчицькою. Було встановлено три Python-бібліотеки: «requests» [1], «beautifulsoup4» [2] і «pandas» [3]. Бібліотека «requests» була використана для надсилання HTTP-запиту до веб-сторінки, що містить список репресованих українських слів та їх сучасних відповідників [1]. Результат запиту

зберігається у вигляді HTML-контенту. Потім, за допомогою бібліотеки «beautifulsoup4» HTML-контент перетворюється у зручну для обробки структуру [2]. Це дозволяє легко знаходити та витягувати необхідні елементи, такі як таблиці та їхні рядки.

Після втановленя Python-бібліотек, використовується метод «find» для знаходження таблиці з відповідними класами, що ідентифікує її як цільову: table = soup.find('table', class_='TblAutoBrdSpc CenteredBlock BTNoIndent') [6]. Далі відбувається витягування заголовків таблиці. З першого рядка таблиці витягуються заголовки, які використовуються як назви колонок у майбутній таблиці. Команда «headers = []» створює порожній список для зберігання заголовків; «table.find_all('td', limit=3)» знаходить перші три теги <td> у таблиці, припускаючи, що це заголовки колонок; «header.text.strip()» витягує текст з тегу <td> і видаляє зайві пробіли з обох боків тексту. Це дозволяє отримати чистий текст заголовків; «headers.append(header.text.strip())» додає очищений текст заголовка до списку «headers». Наступним кроком є витягування рядків таблиці. Особливу увагу приділено обробці спеціальних випадків, таких як рядки з однією колонкою, що є заголовками розділів. Це відбувається за допомогою команди «elif len(cells) == 1:». Останнім кроком є створення DataFrame та збереження у CSV. Дані, зібрані з таблиці, зберігаються у форматі DataFrame за допомогою бібліотеки «pandas» [3]. «pandas» - потужна бібліотека для роботи з даними, аналізу та маніпуляцій з табличними даними (DataFrames) [3]. Цей код забезпечує швидке, точне та автоматизоване імпортування репресованих українських слів та їх сучасних відповідників, що значно полегшує процес створення бази даних для чат-бота.

Автоматично зібрані слова містили багато зайвої інформації, такої як синоніми, посилання на інші слова («див. також») тощо. Наприклад, *«ГОРІШНІЙ штучний архаїзм»* та *«ГОСТРИЦЯ, див. також остриця,*

стята». Така інформація ускладнювала пошук слів та не була необхідною для створення бази даних репресованих слів. Тому вручну було здійснено очищення даних: видалення синонімів, зайвих посилань та іншої несуттєвої інформації. Отже, залишилилось лише *«горішній»* та *«гострица»*. Ручна обробка забезпечила високу якість і точність даних, що використовуються в чат-боті, дозволяючи йому більш точно розуміти і обробляти запити користувачів.

У реєстрі О. В. Демської-Кульчицької всі слова зазначались літерами верхнього регістру [6]. За допомогою фунції «lower» у Google Sheets регістр літер було змінено на нижній.

Наступним важливим кроком став імпорт тлумачень репресованих українських слів та їх сучасних відповідників, що є важливим для повного розуміння їхнього значення та контексту. Спочатку була спроба зробити це вручну, оскільки не всі слова мають тлумачення у сучасних словниках. Однак цей процес виявився дуже трудомістким і часозатратним. Задля цього був розроблений автоматизований скрипт [2], який витягує тлумачення слів із СУМ у 11 томах [29]. Це значно полегшує завдання порівняно з ручним збором даних і забезпечує точність та швидкість отримання необхідної інформації. Пошук тлумачень відбувався на базі онлайн-ресурсу slovnyk.ua [7], в якому містяться тлумачення з словника української мови у 11 томах [29]. Також проводився пошук вручну у словнику української мови у 20 томах [30], для слів, що залишились без тлумачення.

Спочатку імпортуються необхідні бібліотеки: «requests» для відправлення HTTP-запитів і отримання відповідей від веб-сервера [1], «beautifulsoup4» для парсингу HTML та XML документів [2], забезпечуючи зручний інтерфейс для навігації, пошуку і модифікації дерева документа, та «csv» для роботи з CSV-файлами, зокрема для читання і запису даних у табличному форматі [8]. Функція

«get_search_results» відправляє HTTP-запит до онлайн-словника https://slovnyk.ua/index.php, використовуючи параметр запиту «swrd», щоб отримати HTML-контент сторінки результатів пошуку для кожного слова. Функція «parse_meaning» парсить отриманий HTML-контент за допомогою «beautifulsoup4» [2], знаходить та витягує тлумачення слова з відповідних HTML-блоків, зокрема шукаючи блок з класом «box-content» і секції з класом «toggle», де містяться тлумачення. Функція «save_to_csv» зберігає слова і їхні тлумачення у CSV-файл, відкриваючи файл у режимі запису, створюючи об'єкт «csv.writer» і записуючи кожну пару слово-тлумачення у файл. Функція «read_words_from_csv» читає слова з вхідного CSV-файлу, відкриваючи файл у режимі читання, створюючи об'єкт «csv.reader» і зберігаючи кожне слово в список. Основна функція «main» об'єднує всі попередні функції: читає слова з вхідного CSV-файлу, для кожного слова викликає функції «get_search_results» та «parse_meaning», зберігає отримані тлумачення у словник і викликає «save_to_csv» для збереження результатів у вихідний CSV-файл. Таким чином, скрипт забезпечує автоматизоване збирання тлумачень слів [2], що значно полегшує процес створення лінгвістичної бази даних для подальшого використання у розробці чат-бота.

Однак, частину роботи було виконано вручну, оскільки скрипт лише автоматично знаходить тлумачення зі словників, але вибірка потрібних значень і їх подальше оформлення здійснювалось вручну. Це було необхідно, оскільки не всі слова мали тлумачення у сучасних словниках, у таких випадках до слова використовувася термін «не відновлено».

Також зібрана автоматично інформація потребувала редагування через великий обсяг. Наприклад, до слова *«ландшафт»* було знайдено таке тлумачення: *«ЛАНДШ А ФТ , у, ч. 1. Загальний вигляд місцевості ; пейзаж . Макуха попросив льотчика спуститися нижче . Вдивляючись у знайомий ландшафт , став підказувати йому напрямок (Ю. Бедзик,*

Полки .., 1959, 118); Осінній пейзаж змінився одноманітним ландшафтом зими - біла габа снігу вкрила степи (Чаб., Катюша , 1960, 232). 2. Малюнок , картина із зображенням переважно сільської місцевості . Картина на стіні у рамі , під рушником : зимовий ландшафт і вовк на горбі , а внизу у долині засніжене село (Головко , І. 1957, 302). 3. геогр. Частина земної поверхні з певним сполученням рельєфу , клімату , ґрунтів , рослинного і тваринного світу . Типовий ландшафт Нової Зеландії - це дуже гориста місцевість альпійського типу (Посібник з зоогеогр., 1956, 11); В результаті робіт наших експедицій ми прийшли до висновку , що різні природно-вогнищеві хвороби властиві територіям різних географічних ландшафтів (Наука .., 5, 1959, 14).» [29] Для зручності користувачів були видалені синоніми слів, приклади їх використання та деякі значення, що не стосувались слова та його відповідника. Тому в результаті у тлумаченні слова залишилась така інформація: *«1. Загальний вигляд місцевості. 3. геогр. Частина земної поверхні з певним сполученням рельєфу, клімату ґрунтів, рослинного і тваринного світу.»* Це було зроблено з метою виділити лише головну інформацію про слово, оскільки чат-бот спеціалізується саме на репресованих словах та відповідниках до них. Вся інша інформація про ці слова є у відкритому доступі, у онлайн-ресурсах та друкованих виданнях словників. Отже, з метою зосередити увагу користувача на репресованій лексиці вручну була вибрана лише основна інформація про слова [4].

Наступним кроком було написання коду [3], що створює Telegram-бота, який аналізує текстові повідомлення і визначає, чи містять вони репресовані українські слова та їх сучасні відповідники. Розглянемо покроково. Команда «!pip install python-telegram-bot==13.7» використовується для встановлення конкретної версії бібліотеки «python-telegram-bot» (в даному випадку версії 13.7) за допомогою інструменту керування пакетами Python під назвою «pip» [7].

Встановлення конкретної версії пакету «python-telegram-bot» забезпечує сумісність та стабільність коду [7]. Компонент «!» вказує на те, що команда виконується в середовищі Jupyter Notebook для виконання команд оболонки безпосередньо з коду Python. «pip» - це інструмент керування пакетами для Python, який дозволяє встановлювати, оновлювати та видаляти пакети з репозиторію Python Package Index (PyPI). «install» - це команда «pip», яка вказує на необхідність встановлення пакету. «Python-telegram-bot» - це популярна бібліотека для створення Telegram-ботів на Python [7]. До нього додається компонент «==13.7», що вказує на конкретну версію пакету, яку потрібно встановити. У даному випадку версія 13.7.

Наступна команда: «!pip install python-telegram-bot==13.7 pymorphy2 pymorphy2-dicts-uk». Він дозволяє розбирати слова на морфологічні компоненти, наприклад, визначати їх частини мови, основи, закінчення тощо. «Pymorphy2» має кілька переваг над «pymorphy3». Він є більш популярним і широко використовуваним, що забезпечує його стабільність та надійність. «Pymorphy2» має добре задокументовані функції та активну спільноту користувачів, що полегшує отримання підтримки. Він оптимізований для швидкої роботи, мінімального споживання ресурсів та пропонує широкий спектр функцій для морфологічного аналізу, включаючи розпізнавання частин мови, лематизацію та генерацію форм слова. «Pymorphy2-dicts-uk» - це словники для «pymorphy2», які забезпечують підтримку української мови [5]. Вони дозволяють використовувати можливості «pymorphy2» для аналізу українських слів [4]. Комбінація цих пакетів дозволяє створити Telegram-бота, який аналізує українські тексти, обробляє їх морфологічно і надає відповідні відповіді користувачам.

У наступній частині коду спочатку імпортуються необхідні бібліотеки, включаючи «pandas» для роботи з даними [3], «pymorphy2»

для морфологічного аналізу слів та бібліотеки Telegram для створення бота [4]. Встановлюється токен для бота, і завантажується CSV-файл [6], який містить репресовані слова, їх тлумачення, сучасні відповідники та тлумачення цих відповідників. Ініціалізується морфологічний аналізатор «pymorphy2» [4], який генерує всі можливі форми слів. Далі створюється набір значень «word_forms_dict», де ключами є оригінальні слова та їх сучасні відповідники, а значеннями - всі можливі форми цих слів. Такий словник дозволяє боту знаходити слова у будь-яких граматичних формах. Функція «search_word» аналізує вхідне речення, розбиваючи його на окремі слова та словосполучення, та перевіряє наявність цих форм у словнику. Якщо знаходиться відповідність, функція витягує відповідні рядки з DataFrame і формує відповіді, які включають оригінальне репресоване слово, його тлумачення, сучасний відповідник та його тлумачення. Ці відповіді надсилаються користувачеві у форматі HTML. Бот також має обробники команд: «start», який відправляє привітальне повідомлення, і «handle_message», який обробляє текстові повідомлення, викликаючи функцію «search_word» для аналізу тексту. Нарешті, бот запускається у режимі опитування, готовий приймати та обробляти повідомлення від користувачів. Цей підхід дозволяє створити інтерактивного бота, який може допомогти користувачам зрозуміти та використовувати репресовані слова в сучасній українській мові.

У створеного в результаті бота є кілька доступних функцій. Користувач може ввести репресоване слово або словосполучення, і бот надасть тлумачення цього слова, його сучасний відповідник та тлумачення цього відповідника. Наприклад, якщо користувач введе репресоване слово *«безнастанний»*, спочатку бот надасть тлумачення цього слова: *«Який ніколи не припиняється, невідомо, коли закінчиться»*. Також користувач отримає сучасний відповідник - *«неперервний»* разом із його тлумаченням – *«Який триває безперестанно, постійно, без*

перерви». Наступною доступною функцією є пошук репресованого відповідника до слова. Користувач може ввести сучасне українське слово, і якщо до нього є репресований відповідник, бот його знайде та надасть тлумачення обох слів. Наприклад, якщо користувач введе сучасне слово *«грязь»*, бот напише тлумачення: *«Розм'якла від води земля, ґрунт»*, відповідне репресоване слово – *«багно»* і його тлумачення: *«болотисте місце»*. Це допомагає розуміти, як сучасні слова співвідносяться з репресованими термінами. Якщо слово, введене користувачем, використовується лише в сучасній українській мові і не має репресованих відповідників, бот повідомить про це. Така функція забезпечує користувачів інформацією про те, що слово не має історичного контексту репресій, що є важливим для точного розуміння мовних змін.

Крім того бот дає можливість користувачам вводити цілі речення, а не лише окремі слова. Бот аналізує текст, вибирає з нього репресовані слова або сучасні слова, що мають репресовані відповідники, та надає відповідні тлумачення. Наприклад, якщо користувач введе речення *«Сьогодні був сонячний та теплий день, але нараз за вікном з'явилася мряковина»*, бот виявить обидва слова («нараз» і «мряковина»), надасть їх сучасні відповідники та тлумачення. Це робить бот корисним інструментом для глибшого аналізу тексту. Завдяки використанню морфологічного аналізу, бот може розпізнавати різні форми одного і того ж слова, що підвищує точність і коректність відповідей. Наприклад, бот розуміє, що слова *«голодомору»* і «голодомором» є формами одного і того ж слова і відповідно надасть правильну інформацію. Це забезпечує більш детальне розуміння мовних структур та їх змін.

Таким чином, функціональність чат-бота репресованих слів не обмежується простим пошуком значень окремих слів. Він здатен аналізувати текстові повідомлення, визначати репресовані слова та їх сучасні відповідники, надавати їх тлумачення, а також інформувати

користувачів про слова, що використовуються виключно в сучасній мові без історичного контексту цензерування. Це робить бот корисним інструментом для вивчення мовних змін та історії розвитку української мови.

Отже, чат-бот має наступний функціонал:

1. Користувач може ввести репресоване слово, і бот знайде його сучасний відповідник та надасть тлумачення цих слів.
2. Користувач може ввести сучасне слово, і бот знайде відповідне репресоване слово та тлумачення обох слів.
3. Бот може аналізувати введені користувачем речення і виявляти репресовані слова, щоб надати їх сучасні відповідники та тлумачення.
4. У випадку, якщо вхідне повідомлення не містить репресованих слів або їх сучасних відповідників, бот повідомляє користувача про це.

2.3 Тестування ефективності роботи чат-бота

Після розробки чат-бота важливо провести його тестування.

Першим кроком було забезпечення можливості вводити слово як з маленької, так і з великої літери. Це означає, що бот повинен коректно розпізнавати введенний користувачем текст незалежно від регістру літер. Наприклад, введення *«байдужий»* та *«Байдужий»* повинно оброблятися однаково (рис. 2.1, 2.2, 2.3). Тестування на цьому етапі включало перевірку відповідей бота на різні варіанти введення того самого слова,

щоб переконатися, що бот правильно розпізнає і обробляє запити.

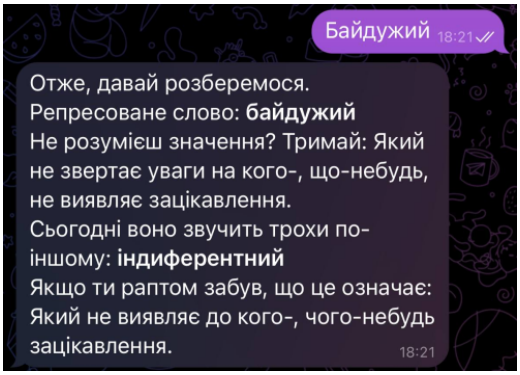


Рис. 2.1 «Байдужий»

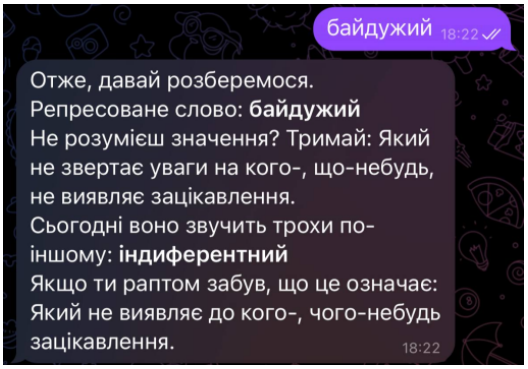


Рис. 2.2 «байдужий»

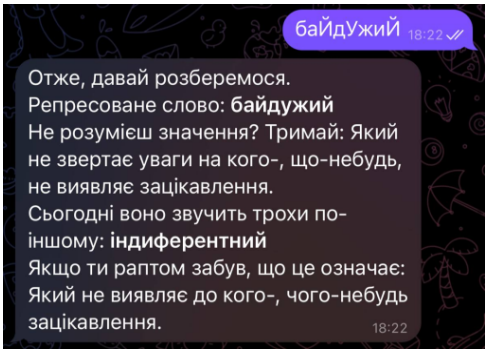


Рис. 2.3 «баЙдУжиЙ»

Після перевірки регістру літер, наступним кроком було тестування бота на здатність правильно обробляти різні слова, що стосуються теми репресованих слів. Перевірялось, чи бот може розпізнавати як репресовані слова, так і сучасні українські слова, що мають репресований відповідник. Наприклад, якщо користувач вводив репресоване слово, бот мав надавати інформацію про його тлумачення і сучасний відповідник (рис. 2.4). Для сучасних українських слів, бот повинен був повідомляти про їх репресовані аналоги, якщо такі існують, та відповідні тлумачення слів (рис. 2.5). Тестування включало перевірку наявності слова в базі даних та відповідне реагування бота, якщо слово відсутнє. Якщо слово було знайдено, бот мав надати відповідну інформацію, а якщо ні – вивести повідомлення про відсутність даних (рис. 2.6).

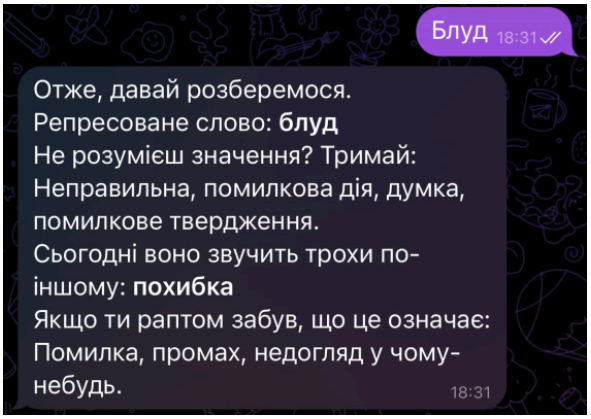


Рис. 2.4 «Блуд»

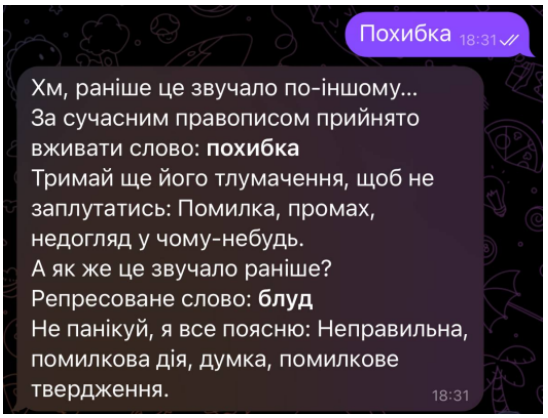


Рис. 2.5 «Похибка»

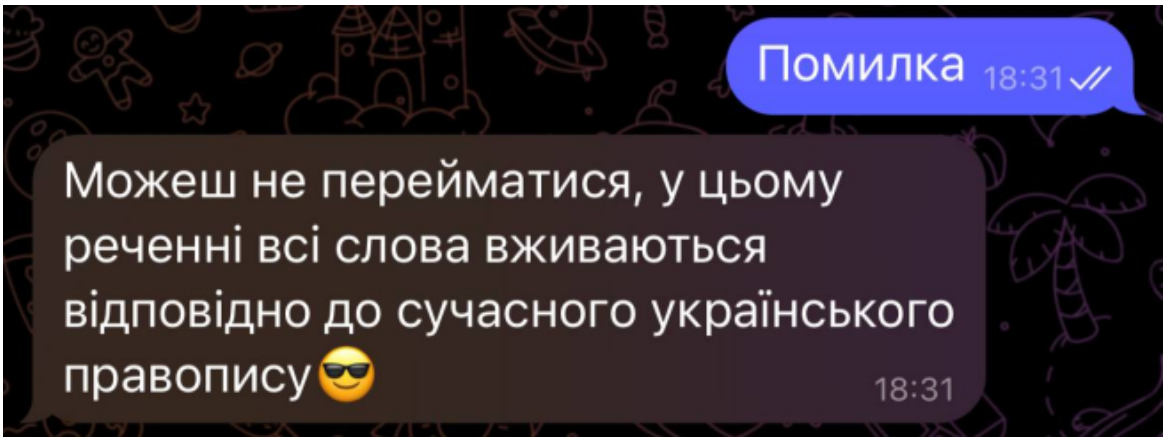


Рис. 2.6. «Помилка»

Наступним кроком було додавання можливості вводити словосполучення. Це вимагало перевірки здатності бота розпізнавати та обробляти запити, що містять кілька слів. Наприклад, бот повинен коректно відповідати на запити типу *«бічна поверхня»* чи *«виробничі сили»*. Тестування включало введення різних комбінацій слів, щоб переконатися, що бот розуміє і правильно обробляє такі запити (рис. 2.7).

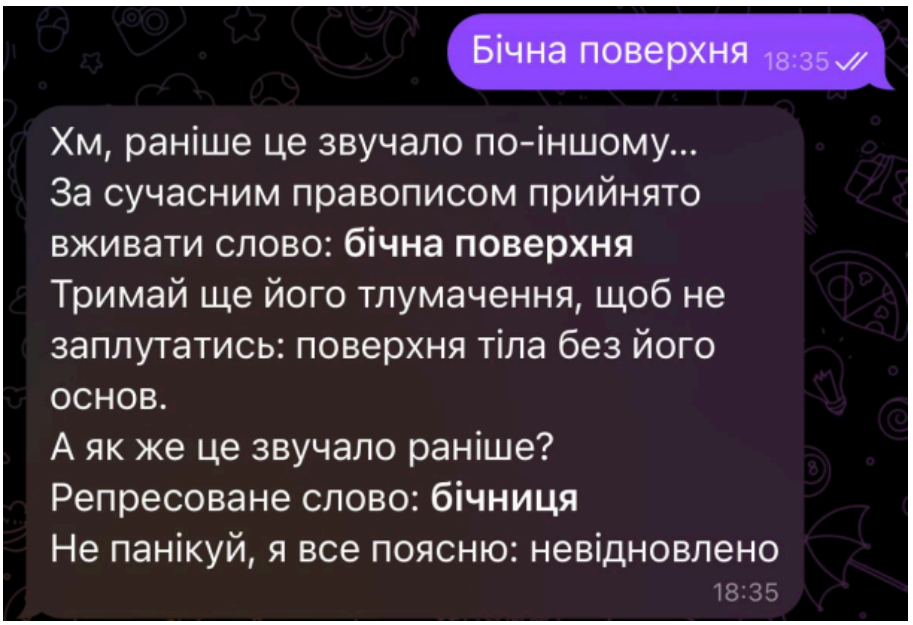


Рис. 2.7 «Бічна поверхня»

Окремим кроком було додавання функції пошуку слів у різних відмінках. Це означає, що бот повинен правильно обробляти та знаходити слова незалежно від їхньої граматичної форми. Наприклад, бот повинен розпізнавати слово *«аркуш»* як *«аркуш»*, *«аркушем»*, *«аркуші»* тощо. Тестування включало введення слів у різних відмінках, щоб переконатися, що бот правильно ідентифікує та обробляє їх (рис. 2.8, 2.9, 2.10).

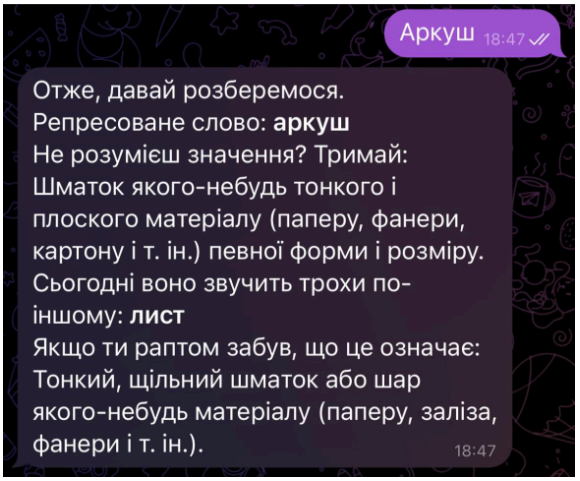


Рис. 2.8 «Аркуш»

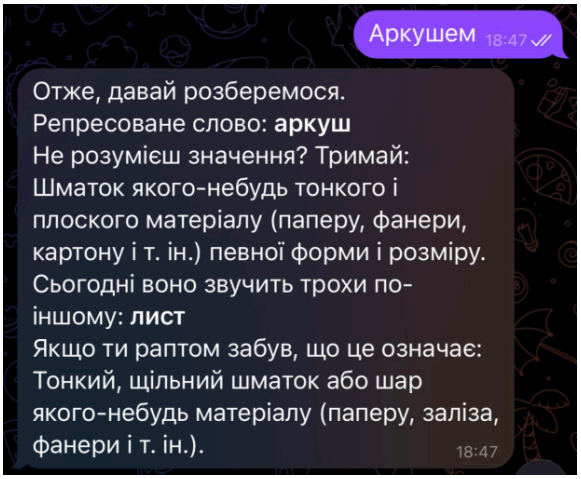


Рис. 2.9 «Аркушем»

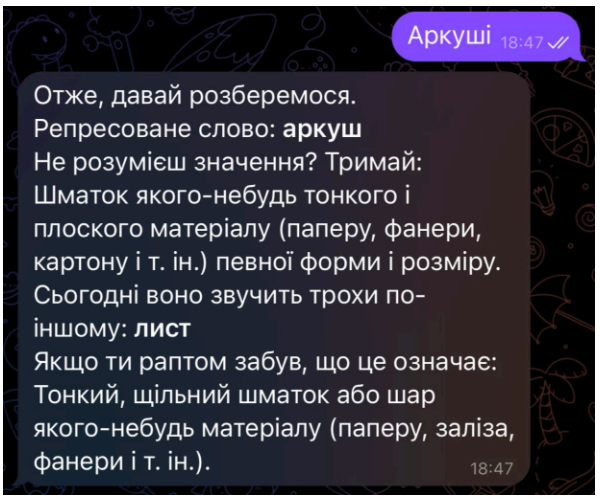
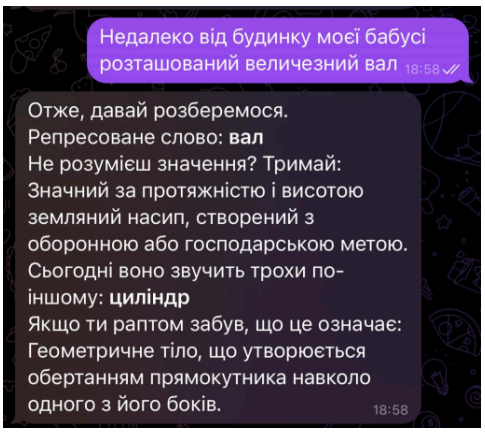


Рис. 2.10 «Аркуші»

Після перевірки регістру літер та здатності обробляти окремі слова та словосполучення, наступним важливим кроком у тестуванні бота було додавання можливості вводити цілі речення. Бот може аналізувати введені користувачами речення, виділяючи з них ключові слова, пов'язані з темою репресованих слів. Наприклад, якщо користувач вводив речення *«На вулиці сьогодні така мряковина»*, бот мав розпізнати слово *«мряковина»* і надати інформацію про його історичний контекст. Оскільки слова можуть зустрічатися у різних відмінках, то бот був налаштований розпізнавати та обробляти слова у всіх можливих формах. Це забезпечило точність відповідей незалежно від граматичної форми слова в реченні. Бот навчався розрізняти репресовані слова та сучасні українські слова, які мають репресований відповідник. Залежно від типу знайденого слова, бот надавав різні відповіді. У випадку, якщо введене слово або слова у реченні не були знайдені в базі даних, бот повідомляє користувача про відсутність інформації.

Процес тестування включав введення різноманітних речень з різними структурами та словами, щоб переконатися у правильності роботи всіх функцій. Було перевірено, чи бот коректно виділяє репресовані слова з речень, розпізнає їх у різних відмінках, правильно

класифікує їх як репресовані або сучасні українські слова з репресованим аналогом і надає відповідні відповіді (рис. 2.11, 2.12).



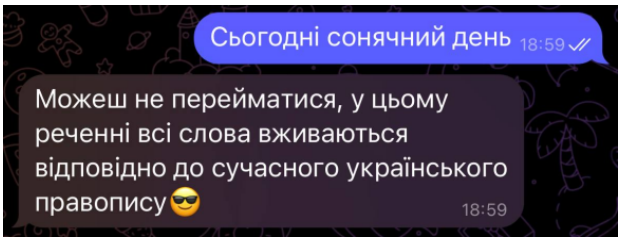


Рис. 2.11 «Речення з репресованим словом» Рис. 2.12 «Речення без репресованих слів»

Таким чином було перевірено ефективність чат-бота і виявлено, що функціонал відповідає поставленим завданням.

Висновки до другого розділу

За результатами дослідження підібрані необхідні технології для розробки чат-бота. Ми обрали створення скриптового типу бота, оскільки такий підхід дозволив обмежити діапазон можливих відповідей до конкретної теми, що забезпечує користувачам зручний та швидкий доступу до інформації про репресовану лексику. Метод обробки даних за визначеними правилами дозволяє попередньо визначити сценарії взаємодії з користувачем і створити набір відповідей на основі цих правил. Це дозволяє користувачам легко знайти необхідну інформацію. Для створення чат-бота було вирішено використовувати платформу Telegram, адже це популярний серед молоді месенджер.

У ході дослідження було розроблено та реалізовано технічне завдання для створення чат-бота, який забезпечує доступ до інформації про репресовану лексику української мови. Виконано послідовність дій, спрямованих на автоматизацію імпорту та обробки даних з реєстру

репресованих слів, створеного О. В. Демською-Кульчицькою, та тлумачень з СУМ у 11 та 20 томах.

Було написано код для автоматичного імпорту репресованих слів та їх сучасних відповідників, очищення даних здійснювалося вручну. Для імпорту тлумачень із СУМу у 11 томах використовувався автоматичний підхід [29], тоді як пошук і редагування тлумачень із СУМу у 20 томах проходило вручну [30].

Розробка чат-бота включала завантаження списку слів із CSV-файлу, встановлення необхідних бібліотек («pandas» [3], «pymorphy2» [4], «python-telegram-bot» [7]) та використання морфологічного аналізатора «pymorphy2» для отримання всіх можливих форм слів [4]. Проведене тестування підтвердило ефективність роботи бота, який має наступний функціонал: введення репресованого слова дозволяє знайти його сучасний відповідник та тлумачення; введення сучасного слова надає відповідне репресоване слово та їх тлумачення; аналіз введених речень для виявлення репресованих слів та надання їх відповідників і тлумачень. Якщо вхідне повідомлення не містить репресованих слів або їх відповідників, бот інформує користувача про це.

Отже, створений чат-бот успішно виконує поставлені завдання, забезпечуючи зручний та швидкий доступ до інформації про репресовану лексику української мови. Використання чат-бота сприяє збереженню та популяризації національної мовної спадщини, роблячи знання про репресовану лексику доступними для широкої аудиторії.

ВИСНОВКИ

У кваліфікаційній роботі «Створення чат-бота «Повертаємо репресовану мову» на платформі месенджера Telegram» представлено алгоритм створення чат-боту репресованої лексики з метою її реактивації [5]. Результати опитування молоді щодо ставлення до репресованої лексики засвідчили актуальність та затребуваність теми дослідження, інтерес серед молоді до історії розвитку української мови. Розглянуто процес реактивації репресованих слів як результат повернення їх до вжитку.

Досліджено Реєстр репресованих слів, укладений О. М. Демською-Кульчицькою, який став основою лінгвістичної бази чат-бота [6].

Результати проведеного дослідження дозволили відповідно до мети та завдань зробити такі висновки та узагальнення:

1) проаналізовано теоретичне підґрунтя явища лінгвоциду, який представлено у контексті свідомого нищення української мови. При цьому встановлено, що поняття «лінгвоцид» розглядається комплексно, у контексті українознавства, тоді як поняття «репресованої лексики» є предметом дослідження саме у сфері лінгвістики.

2) доведено доцільність трактування репресованої лексики як слів, які були насильницько вилучені/замінені з активного вжитку українців радянською владою. А реактивована лексика розуміється як відновлені у мовленнєвій практиці та літературі репресовані слова, які відроджуються з метою національної самототожності українського етносу.

3) проведено лінгвосоціологічне опитування молоді «Ставлення молоді до репресивної мовної політики радянської влади» та констатовано негативне ставлення молоді до цензурування української мови у радянському минулому та зростання інтересу до питомо українських слів. Наслідком репресивної мовної політики вважається

збіднення української мови, штучне зближення української та російської мов, розвиток двомовності.

4) було створено лінгвістичну базу даних репресованих слів [4], шляхом їхнього імпорту з Реєстру репресованих слів, укладеного О. М. Демською-Кульчицькою. А також імпорт тлумачень з СУМ у 11 томах [29] та СУМ у 20 томах [30]. Необхідна інформація була зібрана за допомогою веб-скрапінгу [1, 2];

5) створено код для автоматичної роботи чат-бота на платформі Telegram [3];

6) протестовано чат-бот репресованої лексики на платформі Telegram як інтерактивний ресурс для користувачів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Агеєва В. П. За лаштунками імперії. Київ : Віхола, 2021. 360 с.
2. Антонів Ю. В. До проблеми реабілітації лексики, вилученої з української мови. *Наукові записки.* Том 20. Філологічні науки. С.38-42.
3. Бондар О. І. Екологія українського слова: аспекти і проблеми // *Мовознавство*: Доп. та повідомл. IV Міжнар. конгр. україністів / Відп. ред. В. В. Німчук. К.: Пульсари, 2002. С. 158-163.
4. Васильєв В. Ю. Український правопис. Енциклопедія історії України: Україна-Українці. Кн. 2 / Редкол. : В. А. Смолій (голова) та ін. НАН України. Інститут історії України. К. : В-во «Наукова думка», 2019. 842 с.. URL: http://www.history.org.ua/?termin=ukrajinskyj_pravopys_1928
5. Гавдида Н. І., Назаревич Л. Т. Лінгвоцид як форма мовної політики. Наукові записки ТНПУ. Серія : Мовознавство. 2014. Вип. II (24). С. 77–80. URL: http://dspace.tnpu.edu.ua/bitstream/123456789/6456/1/Havdyda.pdf
6. Демська-Кульчицька О. М. Реєстр репресованих слів. *Українська мова у XX сторіччі : історія лінгвоциту : Док. і матеріали* / Упоряд. : Л. Масенко та ін. К. : Вид. дім «Києво-Могилянська академія», 2005. С. 354-399.
7. Історія українського правопису XVI–XX століття: Хрестоматія / Упорядн.: В. В. Німчук, Н. В. Пуряєва. К.: Наукова думка, 2004. 582 с. URL: http://history.org.ua/LiberUA/966-00-0261-0/966-00-0261-0.pdf
8. Караванський С. Й. Секрети української мови: Науково-популярна розвідка з додатком словничків репресованої та занедбаної української лексики. Київ : УКСП «Кобза», 1994. 152 с.

9. Караванський С. Й. Пошук українського слова, або боротьба за національне «я». Київ : Видавничий центр «Академія», 2001. 240 с.
10. Качуровський І. 8th Доля українських лінгвістів на тлі радянської мовної політики. *Репресовані мовознавці* : збірник наукових праць / науковий редактор Бойко Н. І. Ніжин : Видавництво НДУ ім. М. Гоголя, 2010. 103 с.
11. Костомаров М. І. Дві руські народності / Микола Костомаров ; [упоряд.: І. Сюндюков, М. Томак, Н. Тисячна ; за заг. ред. Л. Івшиної]. К. : Укр. прес-група, 2012. 71 с.
12. Лизанчук В. В. Геноцид, етноцид, лінгвоцид української нації: хроніка. Львів : Видавничий центр Львівського національного університету імені Івана Франка, 2008. 258 с.
13. Мазурик Д. В. Сучасні тенденції в оновленні лексики української літературної мови. *Вісник Львівського університету*. Серія філологічна. Т. 29, 2000. С. 177-182.
14. Масенко Л. Т. Мова і суспільство: Постколоніальний вимір. Київ : Вид. дім «Києво-Могилянська академія», 2004. 164 с.
15. Масенко Л. Т. Стратегія дослідження мовної ситуації в Україні та мовного розвитку. *Українська мова.* № 4 (84), 2022. С. 8.
16. Масенко Л. Т. Радянська спадщина в мовній політиці України. *Екологія мови і мовна політика в сучасному суспільстві* / за ред. У Б. М. Ажнюк. Київ: Видавничий дім Дмитра Бураго, 2012. С 65–72.
17. Мацько Л. І. Лінгвосоціокультурний аспект «Словника української мови-20» *Система і структура східнослов'янських мов.* 2012. Вип. 5. С. 200-209. URL: http://nbuv.gov.ua/UJRN/sissm_2012_5_31
18. Огієнко І. Українська культура / І. Огієнко. К., 1918. 273 с.
19. Паламарчук Л. С. Словник української мови (СУМ) / НАН України, Інститут мовознавства ім. О. О. Потебні, Інститут української мови; ред. В. М. Русанівський [та ін.]. К. : Українська енциклопедія, 2000.

20. Поздрань Ю. В. «Російсько-український словник» за редакцією А. Ю. Кримського та С. О. Єфремова в історико-лінгвістичному контексті : монографія / за наук. ред. О. М. Тищенко. Вінниця, 2018. 292 с. URL: http://ir.lib.vntu.edu.ua//handle/123456789/23269
21. Провотар О.І. Особливості та проблеми віртуального спілкування за допомогою чат-ботів / О. І. Провотар, Х. А. Клочко. Наукові праці ВНТУ: Інформаційні технології та комп'ютерна техніка. 2013. № 3. 6 с.
22. Радевич-Винницький Я. К. Лінгвоцид як форма геноциду. Київ : Українська видавнича спілка ім. Ю. Липи, 2011. 78 с.
23. Радевич-Винницький Я. К. , Іванишин В. П. Мова і нація : тези про місце і роль мови в національному відродженні України. Дрогобич : Відродження, 1994. 217 с.
24. Ренчка І. Є. Лексикон тоталітаризму. Київ : КЛІО, 2018. 232 с.
25. Реєстр репресованих слів. Мислене древо. URL: https://www.myslenedrevo.com.ua/uk/Sci/Linguistics/rejestr.html
26. Рудницький Я. Б. Етимологічний словник української мови у 2-х т. Вінніпег-Оттава, 1962. 82 с.
27. Сваричевська А. П. Лінгвоцид або ж лінгвістичний геноцид української мови. Нова філологія, 2023. URL: http://zfs-journal.uzhnu.uz.ua/archive/27/part_3/1.pdf
28. Слісаренко М. Чат-бот : поняття, історія розвитку, перспективи застосування. Науково-дослідна робота студентів як чинник удосконалення професійної підготовки майбутнього вчителя : зб. наук. пр. / Харків. нац. пед. ун-т ім. Г. С. Сковороди ; редкол. : Н. О. Пономарьова, Н. В. Оліфіренко, В. М. Андрієвська та ін. Харків, 2024. Вип. 23. С. 124-133.
29. Словник української мови в 11 томах. URL: https://sum.in.ua/

30. Словник української мови: у 20 томах. / НАН України, Укр. мов.-інформ. фонд. Київ : Наукова думка, 2010. URL: https://sum20ua.com/?wordid=0&page=0
31. Українська мова у XX сторіччі: історія лінгвоциду : документи і матеріали / Всеукр. т-во «Просвіта» ім. Т. Шевченка; за ред. Л. Масенко ; упоряд. Л. Масенко [та ін.] . Київ : Києво-Могилянська академія, 2005. 399 с.
32. Український правопис / Нар. комісаріят освіти УСРР. Вид. 1-ше. Київ : Держ. вид-во України, 1929. 103 с.
33. Тулузакова О. Г. Актуалізація «репресованої» лексики (на матеріалі творів представників «Станіславського феномену»). *Наукові записки*. Том 85. Філологічні науки. 2008. С. 55-63.
34. Тулузакова О. Г. Актуалізація лексики: стан дослідження проблеми. *Наукові праці:* Серія «Філологія. Мовознавство». Чорном. держ. ун-т ім. Петра Могили. Миколаїв, 2016. Вип. 243. Том 255. С. 92–95. URL: https://www.researchgate.net/publication/321807891_AKTUALIZACIA_LEKSIKI_STAN_DOSLIDZENNA_PROBLEMI
35. Франко І. Двоязичність і дволичність // Франко І. Мозаїка. Львів, 2001. С. 263-277.
36. Шевельов Ю. В. Українська мова в першій половині двадцятого сторіччя (1900–1941). Стан і статус / Ю. В. Шевельов. Чернівці : Рута, 1998. 208 с.
37. Шишкіна В. О. Створення інформаційного чат-боту для студентів ТНТУ засобами мови програмування Python та Telegram API: кваліфікаційна робота освітнього рівня «Бакалавр» «122 - комп'ютерні науки» / В. О. Шишкіна. Тернопіль : ТНТУ, 2022. 44 с.

38. Andre E, Pelachaud C. Interacting with Embodied Conversational Agents. URL: https://www.researchgate.net/publication/225995164_Interacting_with_Embodied_Conversational_Agents
39. Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, Agathe Merceron (Eds.). Handbook of Learning Analytics (2nd. ed.). SoLAR, Vancouver, 2022. 244 p. URL: https://solaresearch.org/wp-content/uploads/hla22/HLA22.pdf
40. Chayinska M., Kende A., Wohl M. J. A. National identity and beliefs about historical linguicide are associated with support for exclusive language policies among the Ukrainian linguistic majority. Group Processes & Intergroup Relations. 2022. Vol. 25 (4). P. 924–940.
41. Sumit Raj Building Chatbots with Python: Using Natural Language Processing and Machine Learning. URL: https://www.academia.edu/40419686/Building_Chatbots_with_Python_Using_Natural_Language_Processing_and_Machine_Learning_Sumit_Raj?auto=download
42. Shawar Bayan Abu, Atwell Eric. Chatbots: Are they Really Useful? URL: https://www.researchgate.net/publication/220046725_Chatbots_Are_they_Really_Useful
43. Romao Gil A Literature Review on chatbots in education: An intelligent chat agent. URL: https://www.researchgate.net/publication/352384737_A_Literature_Review_on_chatbots_in_education_An_intelligent_chat_agent
44. Hussain Shafquat, Sianaki Omid Ameri, Ababneh Nedal A Survey on Conversational Agents/Chatbots Classification and Design Techniques. URL:

https://www.researchgate.net/publication/331746678_A_Survey_on_Conversational_AgentsChatbots_Classification_and_Design_Techniques

45. Telegram raises $210 million through bond sales. URL: https://techcrunch.com/2023/07/18/telegram-raises-210-million-through-bond-sales/

Інтернет-джерела програмних бібліотек:

1. requests [Електронний ресурс]. — Режим доступу : https://pypi.org/project/requests/
2. beautifulsoup4 [Електронний ресурс]. — Режим доступу : https://pypi.org/project/beautifulsoup4/
3. pandas [Електронний ресурс]. — Режим доступу : https://pypi.org/project/pandas/
4. pymorphy2 [Електронний ресурс]. — Режим доступу : https://pypi.org/project/pymorphy2/
5. pymorphy2-dicts-uk [Електронний ресурс]. — Режим доступу : https://pypi.org/project/pymorphy2-dicts-uk/
6. метод «find» [Електронний ресурс]. — Режим доступу : https://pypi.org/project/find/
7. python-telegram-bot [Електронний ресурс]. — Режим доступу : https://pypi.org/project/python-telegram-bot/
8. csv [Електронний ресурс]. — Режим доступу : https://docs.python.org/uk/3/library/csv.html
9. Google Colab [Електронний ресурс]. — Режим доступу : https://colab.research.google.com/?hl=uk

Додаток 1

Опитувальник

«Ставлення молоді до репресивної мовної політики радянської влади»

Вступна частина. В історії розвитку української мови відоме явище обмеження радянською владою користування слів, що вживалися до 1930-років. Багато українських слів було вилучено чи замінено на російський відповідник, який «вживляли» у словники та, відповідно, який ставав літературною нормою.

Ваші відповіді допоможуть з'ясувати вплив репресованих українських слів на національне самовизначення молоді.

Опитування анонімне, результати будуть використані в узагальненому вигляді.

ДЯКУЄМО за співпрацю!

Основна частина.

1. Якою мовою ви спілкуєтесь у повсякденному житті?
    - українською мовою;
    - російською мовою;
    - українською та російською;
    - іншою.
2. Як ви оцінюєте свій рівень володіння українською мовою за 10-бальною шкалою, де 1 бал - початковий рівень, а 10 - високий.
3. Чи відомо вам про факт репресивної мовної політики радянської влади в Україні?
    - так;
    - ні.

4. Чи знаєте ви які саме слова було вилучено/замінено внаслідок політики "зрощення" української та російської мов?
    - так;
    - частково;
    - ні.
5. На вашу думку, що означає слово "підсоння"?
    - клімат
    - підвіконня
    - під сонцем
6. На вашу думку, що означає слово "оник"?
    - нуль
    - онук
    - коник
7. На вашу думку, чи існує мовна проблема в Україні?
    - так;
    - ні.
8. На вашу думку, які наслідки мала така мовна політика?
    - Збіднення української мови та її регресія.
    - Збагачення української мови та поповнення словникового запасу.
    - Штучне зближення української та російської мов.
    - Задоволення потреби у нових словах індустріального суспільства.
    - Розвиток двомовності.
    - Це надумана проблем.
9. Оцініть за 10-бальною шкалою вплив репресивної мовної політики радянської влади на мовну ситуацію, де 1 бал - незначний, а 10 - дуже вагомий.
10. Як би ви зреагували на вживання репресованих слів у медіа?

- позитивно;
- негативно;
- байдуже.

11. Чи згодні ви з твердженням, що найбільш спорідненою та близькою до української мови є російська?
    - так;
    - ні.

Додаток 2

1. Скрипт для веб-скрапінгу репресованих слів та їх відповідників. URL: https://colab.research.google.com/drive/1zWlJwJ8zej-kxZ1j6bSfRRpT8KwsCA_6#scrollTo=81SeW1mlqzn0
2. Скрипт для веб-скрапінгу тлумачень. URL: https://colab.research.google.com/drive/174Rm-JjXO_tMPm-xceAREdaRpwAo3CfK
3. Скрипт для створення чат-боту. URL: https://colab.research.google.com/drive/1nQ8RgOREM58wDeBfRFQ9RTcS_Nff08lv#scrollTo=o2iAEOgwjUrc
4. Створена база даних репресованих слів, їх відповідників та тлумачень цих слів. URL: https://docs.google.com/spreadsheets/d/1jHlRt1kirsgokn-U3zJLZnZhHkmX6MMYK4wKiNgiSow/edit#gid=0
5. Створений чат-бот. URL: https://t.me/repressed_words_bot

6. Папка на гугл-диску з усіма файлами. URL: https://drive.google.com/drive/u/0/folders/1lRNqJWBsB7VtemUz58iZU50ZsplNOFKK
7. Онлайн-ресурсу slovnyk.ua. URL: https://slovnyk.ua/
8. Опитування «Ставлення молоді до репресивної мовної політики радянської влади»: https://forms.gle/szx6a3FgBzzALKVu7