Знайомимо з корпусом СВІДОК та словником.

Інформаційно-документальний потенціал корпусу текстів СВІДОК (**С**учасний **в**оєнний **і**нформаційно-**д**окументальний **о**писовий **к**орпус) (https://korpusomat.eu/corpus/6625)

Повномасштабна збройна агресія рф проти України спричинила глобальні зрушення не лише в геополітичному та соціальному вимірах, а й у мовній дійсності. Збройний конфлікт сформував унікальний за своєю інтенсивністю, динамікою та емоційним навантаженням дискурс. Документування, фіксування та систематизування цього текстового масиву є критично важливим завданням як для збереження його як української культурної спадщини, історичної пам'яті, так і для здійснення різноаспектних лінгвістичних, соціогуманітарних та прикладних міждисциплінарних досліджень.

Відповіддю на цей науковий виклик став **СВІДОК** – Сучасний воєнний інформаційно-документальний описовий корпус.

Мета та специфіка корпусу

СВІДОК розроблено як спеціалізований інформаційно-документальний ресурс, покликаний репрезентувати зріз сучасного українського воєнного дискурсу. Основна мета створення корпусу – надати науковій спільноті репрезентативний, систематизований та верифікований масив текстових даних, що відображають мовну та соціокультурну картину світу українського суспільства в умовах війни.

Основними джерелами формування корпусу слугують:

офіційні та медійні джерела, тобто воєнні зведення, офіційні звернення, аналітичні та репортажні матеріали медіа;

его-документи та наративи свідчень, тобто спогади, щоденники, інтерв'ю цивільних і військових, що фіксують індивідуальний та колективний досвід переживання травми;

літературно-художній дискурс, тобто літературні твори сучасної воєнної доби;

цифровий/мережевий дискурс, тобто публікації із соціальних мереж та месенджерів (Telegram, Facebook тощо), які відображають спонтанну мовну/мовленнєву реакцію суспільства в режимі реального часу.

Наукова цінність та методологійні перспективи

Унікальність корпусу **СВІДОК** полягає в поєднанні документальної достовірності з описовим та аналітичним потенціалом. Використання корпусної методології вможливлює перехід від описового спостереження за мовними змінами до верифікованого аналізу великих масивів даних (Big Data).

Завдяки спеціальній структурній та лінгвістичній розмітці (анотації) корпус дає змогу досліджувати:

1) лексико-семантичні й неологійні процеси, як-от: виникнення воєнних неологізмів, трансформацію семантики звичайних слів (трансформація конотацій), функціонування метафор та евфемізмів;

2) дискурс-аналіз та когнітивні студії, тобто аналіз концептуальних метафор війни, засобів вираження травми, концептуалізації понять «свій/чужий», ворожого та патріотичного дискурсів;

3) психолінгвістичні та соціолінгвістичні параметри, зокрема динаміку мовної поведінки суспільства під час стресу, трансформацію мовного етикету та оцінної лексики;

4) застосування корпусних даних у NLP (Natural Language Processing), тобто використання даних для навчання моделей штучного інтелекту, систем автоматичного аналізу тональності (sentiment analysis) та вилучення сутностей (NER) у сфері специфічної воєнної термінології.

Отже, корпус текстів **СВІДОК** постає не лише методологійним інструментом для мовознавців, соціологів, істориків і фахівців з комп'ютерної лінгвістики, але й цифровою пам'яткою сучасної української стійкості. Систематизоване збереження мовного матеріалу війни створює надійне підґрунтя для майбутніх фундаментальних розробок, забезпечуючи об'єктивність досліджень через прямий доступ до першоджерел.

Фінансування та підтримка

СВІДОК – це проєкт колектива українських науковців у межах дербюджетної теми МОН України «Актуальні системні зміни в українській мові воєнної доби на тлі загальноєвропейських лінгвальних трансформацій» (2-697-26), який є підкорпусом Генерального регіонально анотованого корпусу української мови (ГРАК).

ГРАК – це волонтерський проєкт, розміщений на базі Єнського університету. Окремі технічні роботи та частину робіт з підготовки колекції газетних текстів профінансовано за рахунок грантів Єнського університету (2019–2024 рр.).

Просимо посилатися на ГРАК та СВІДОК

Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, Т. Ніколаєнко, А. Лукашевський та ін. — Київ, Львів, Єна, 2017–. — uacorpus.org.

Сучасний воєнний інформаційно-документальний описовий корпус (СВІДОК) / І. Попова, Н. Дьячок, О. Гурко, О. Куварова, О. Панченко, Ж. Тернова. – Дніпро, 2026–. — uacorpus.org.

Мовна динаміка воєнного часу: лексикографічний та документальний вимір словника «СВІДОК» (https://www.lexonomy.eu/dictionary/svidok/home)

Повномасштабна збройна агресія рф проти України спричинила кардинальні зрушення в усіх сферах життєдіяльності українського суспільства, що неминуче знайшло відображення в мовній системі. Мова як чутливий індикатор соціальних змін першою чергою реагує на екстремальні події, акумулюючи нові психоемоційні, соціокультурні та поняттєво-термінологійні смисли. Сучасний український воєнний дискурс продемонстрував безпрецедентну динаміку лексико-семантичного оновлення, сформувавши потужний пласт актуальної лексики та новотворів (неологізмів, потенційних слів, оказіоналізмів).

Словник **СВІДОК** (*Сучасний воєнний інформаційно-документальний описовий корпус*) постає результатом наукового осмислення, систематизації та лексикографічного опрацювання цього мовного масиву.

Концептуальні засади та джерельна база

Основна мета електронного словника – зафіксувати й описати лексико-семантичний стан української мови воєнної доби, репрезентувавши як вузькоспеціалізовані термінологійні інновації, так і зафіксовані в загальновживаному мовленні неологізми, метафори, евфемізми та словотвірні модифікації.

Джерельною базою для укладання словника слугували матеріали електронного корпусу **СВІДОК**, що охоплює різножанрові текстуальні масиви.

Медійний та офіційно-інформаційний дискурс: воєнно-аналітичні зведення, офіційні звернення, репортажі вітчизняних та зарубіжних ЗМІ;

Его-документи й естетично-художні тексти: щоденники свідків подій, спогади, інтерв'ю, художня та публіцистична проза;

Мережевий дискурс (SoMe/TG): публікації та коментарі в соціальних мережах і месенджерах, що відображають живий, спонтанний мовний потік і масову словотворчість.

Лексико-семантичні вектори новотворів

Подана в словнику лексика відображає ключові тенденції словоскладання, семантичної деривації та запозичень, серед яких виокремлені такі групи.

1. Спеціальна та військово-технічна термінологія: запозичення і адаптація назв новітніх озброєнь, тактичних прийомів та засобів ведення війни, які набули масового вжитку.

2. Семантична переорієнтація та реконцептуалізація: трансформація конотацій та розширення значення вже наявних у мові лексем під впливом воєнного контексту.

3. Експресивно-оцінна та іронічна лексика: метафоричні новотвори, засоби лінгвістичного опору, сатиричні та меметичні одиниці, що виконують психотерапевтичну та консолідуючу функції в суспільстві.

4. Соціолінгвістичні та психолінгвістичні маркери: лексика на позначення нових соціальних ролей, статусів та психологійних станів людини в умовах війни.

Практична та наукова цінність

Укладання словника **СВІДОК** базується на принципах синхронного опису з урахуванням документального верифікування кожної лексичної одиниці. Ілюстративний матеріал словникових статей унаочнює реальне функціонування слів у контексті, зафіксованому зокрема в корпусі, що дає змогу простежити їхню сполучуваність та стилістичне забарвлення.

Словник орієнтований на широке коло фахівців: мовознавців, соціолінгвістів, дискурс-аналітиків, істориків, медіадослідників, перекладачів, а також усіх, хто цікавиться процесами розвитку сучасної української мови. Окрім суто наукового значення, видання виконує важливу культурно-меморіальну функцію, зберігаючи мовний портрет епохи та закарбовуючи досвід українського спротиву для майбутніх поколінь.

ПРАВИЛА КОРИСТУВАННЯ ЕЛЕКТРОННИМ СЛОВНИКОМ «СВІДОК»

Електронний словник воєнних новотворів та актуальної лексики **СВІДОК** розроблено як інтерактивний науково-довідковий ресурс. Словник поєднує класичні лексикографічні традиції з можливостями сучасних інформаційних технологій (корпусного аналізу, автоматизованого пошуку та динамічного фільтрування даних).

1. Структура та елементи словникової статті

Кожна електронна словникова стаття має уніфіковану структуру та складається з таких обов’язкових і додаткових блоків.

Заголовне слово (реєстрове слово / за Used-term): подано за абеткою в початковій формі (для іменників — називний відмінок однини; для дієслів — інфінітив; для прикметників — чоловічий рід, однина, називний відмінок). Наголос позначено графічно.

Граматична характеристика: подана як скорочені позначки одразу після заголовного слова (наприклад: *ім., ж. р., розм.*, *дієсл., недок. в.*).

Стилістичні та функційні позначки (маркери): вказують на сферу вживання, емоційно-експресивне забарвлення або походження лексеми (наприклад: *військ.*, *неол.*, *жарг.*, *ірон.*, *мереж.*, *іншомовн.*).

Семантичний блок (тлумачення): розкриває значення лексичної одиниці, особливості її контекстуального функціонування або переносне значення.

Етимологійна / деривацийна довідка (за наявності): містить відомості про походження слова, твірну основу, джерело запозичення чи модель словотвору.

Ілюстративний матеріал (ілюстративні текстові контексти): приклади вживання лексеми, автоматично або напівавтоматично дібрані з корпусів **СВІДОК, ГРАК** або інших електронних ресурсів. Кожен контекст постачається метаданими:

Джерело (офіційне зведення, медіа, соцмережі, его-документ);

Дата фіксації (ДД.ММ.РРРР);

Гіперпосилання на першоджерело (за наявності та умов збереження конфіденційності).

Блок зв’язаних одиниць (Cross-references): перелік синонімів, антонімів, гіпонімів, гіперонімів та пов'язаних концептів у формі активних гіперпосилань.

2. Функціонал пошуку та навігації

Для ефективного пошуку інформації в інтерфейсі електронного словника передбачено кілька режимів:

2.1. Простий та розширений пошук

Простий пошук: здійснюється через пошуковий рядок за повним словом або його частиною.

Розширений пошук: дозволяє комбінувати параметри:

Пошук за *початковою формою* або за *довільною словоформою* (морфологічний пошук);

Пошук у *тлумаченнях* (пошук за ключовими словами в тексті визначень);

Пошук за *ілюстративними контекстами* (цитатами з корпусу).

2.2. Система динамічної фільтрації (Тегування)

Користувач може відфільтрувати словникові статті за такими параметрами.

За частинами мови (*іменники, дієслова, прикметники тощо*).

За тематичними групами / концептами (*«Озброєння та техніка», «Психоемоційні стани», «Ворожа пропаганда та протидія», «Соціальні ролі» тощо*).

За хронологією (фільтрація за роками та місяцями першої фіксації слова в корпусі).

За типом дискурсу (*офіційний, медійний, мережевий, художній*).

2.3. Алфавітний покажчик

Двомовний (українсько-англійський для іншомовних запозичень) інтерактивний алфавітний покажчик дає змогу здійснювати швидкий перехід до потрібної літери або словосполучення.

3. Прийняті скорочення та умовні позначки

3.1. Граматичні скорочення

Скорочення	**Повне значення**
ім.	іменник
прикм.	прикметник
дієсл.	дієслово
присл.	прислівник
чол. р. (ч.)	чоловічий рід
жін. р. (ж.)	жіночий рід
сер. р. (с.)	середній рід
одн. / **мн.**	однина / множина
док. в. / **недок. в.**	доконаний вид / недоконаний вид

3.2. Стилістичні та функційно-термінологійні позначки

Позначка	Значення / Сфера застосування
військ.	військова та військово-технічна термінологія
неол.	неологізм (нове лексичне утворення воєнної доби)
оказ.	оказіоналізм (авторське або контекстуальне слово)
мереж.	мережевий жаргон / сленг соціальних мереж та месенджерів
розм.	розмовне слово або вираз
жарг.	жаргонізм
ірон.	іронічне вживання / сатиричний контекст
перен.	переносне значення
оцін.	лексика з вираженою позитивною чи негативною оцінкою

4. Інтеграція з корпусом та цитування

1. Динамічний зв'язок з корпусом: натискання на кнопку **«Переглянути в корпусі»** під словниковою статтею перенаправляє користувача до модуля корпусного пошуку **СВІДОК**, де подано повний перелік усіх конкордансів (усіх випадків уживання певного слова в базі даних).

2. Академічне цитування: у верхньому правому кутку кожної словникової статті розміщено кнопку **«Цитувати»**, яка автоматично генерує посилання на конкретну словникову статтю у форматах APA, MLA, Harvard та ДСТУ 8302:2015.

Фінансування та підтримка

СВІДОК – це проєкт колектива українських науковців у межах дербюджетної теми МОН України «Актуальні системні зміни в українській мові воєнної доби на тлі загальноєвропейських лінгвальних трансформацій» (2-697-26), який є структурним складником Генерального регіонально анотованого корпусу української мови (ГРАК).

ГРАК – це волонтерський проєкт, розміщений на базі Єнського університету. Окремі технічні роботи та частину робіт з підготовки колекції газетних текстів профінансовано за рахунок грантів Єнського університету (2019–2024 рр.).

Просимо посилатися на ГРАК та СВІДОК

Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, Т. Ніколаєнко, А. Лукашевський та ін. — Київ, Львів, Єна, 2017–. — uacorpus.org.

Сучасний воєнний інформаційно-документальний описовий корпус (СВІДОК) / І. Попова, Н. Дьячок, О. Гурко, О. Куварова, О. Панченко, Ж. Тернова. – Дніпро, 2026–. — uacorpus.org.