

<https://doi.org/10.31891/2307-5732-2026-363-24>

УДК 004.912:004.8

САВИЦЬКИЙ РОМАН

Державний університет “Житомирська політехніка”

<https://orcid.org/0000-0003-3617-7900>

Email: roman.savitskyi@gmail.com

ВАКАЛЮК ТЕТЯНА

Державний університет “Житомирська політехніка”

<https://orcid.org/0000-0001-6825-4697>

Email: tetianavakaliuk@ztu.edu.ua

ГІБРИДНА СИСТЕМА ГЕНЕРАЦІЇ ЛЕКСИЧНИХ ЗАМІН УКРАЇНОМОВНОГО ТЕКСТУ ДЛЯ ПІДВИЩЕННЯ КОГНІТИВНОЇ ДОСТУПНОСТІ

У статті розглянуто генерацію заміन на основі словника та LLM для спрощення українських текстів із фокусом на підвищенні когнітивної доступності. Запропоновано архітектуру, де словниковий пошук є першим рівнем, а мовні моделі використовуються як наступний рівень пошуку для невідомих або складних одиниць. Результати фільтруються, кешуються та повертаються в текст із морфологічним узгодженням. Описано сценарії використання для освітніх, технічних та інструктивних текстів. Проведено порівняння чотирьох моделей, що використовуються або є цільовими в проекті: *MamaLLM*, *LapaLLM*, *Gemma 3 12B* і *Gemini 1.5 Flash*. Оцінювання виконано за чотирма групами метрик, а саме якість замін, швидкодія, покриття та ресурсна вартість. Результати підтверджують доцільність гібридного підходу. Словниковий рівень забезпечує стабільність і швидкість, LLM підвищують семантичну адекватність для складних категорій, а кешування знижує кількість повторних запитів. Сформульовано практичні рекомендації щодо вибору моделей і сценаріїв застосування в системах спрощення українських текстів.

Ключові слова: спрощення тексту; лексична заміна; словник; LLM; когнітивна доступність; українська мова

SAVITSKYI ROMAN, VAKALIUK TETIANA

Zhytomyr Polytechnic State University

HYBRID SYSTEM FOR GENERATING LEXICAL SUBSTITUTIONS IN UKRAINIAN TEXT TO ENHANCE COGNITIVE ACCESSIBILITY

The paper presents a substitution generation approach based on a dictionary and an LLM for simplifying Ukrainian texts, with the goal of improving cognitive accessibility. The proposed architecture is organized as a layered pipeline. A dictionary lookup provides fast and stable replacements for known lexical items, while large language models are used as a fallback for rare or complex cases. Candidate replacements are filtered, cached, and reused, and the final output is morphologically aligned with the source context. We describe practical usage scenarios in which the system demonstrates optimal utility, including educational materials, technical documentation, and step-by-step instructions. An experimental comparison is carried out across four models used or targeted in the project, namely *MamaLLM*, *LapaLLM*, *Gemma 3 12B*, and *Gemini 1.5 Flash*. Evaluation is structured around four metric groups: substitution quality, response speed, coverage as the share of cases with at least two valid alternatives, and resource cost. The results highlight typical trade-offs. Larger models provide better semantic adequacy for complex categories, while smaller or faster models are more suitable for lightweight lexical substitutions and real-time usage. The dictionary-first strategy stabilizes output quality and reduces latency, whereas the caching mechanism minimizes repeated LLM requests and supports scalable deployment. The study consolidates an operational design for hybrid simplification, documents the role allocation between dictionaries and LLMs, and provides recommendations for selecting model pairs depending on category complexity and deployment constraints. The findings are directly grounded in the current project infrastructure and offer a reproducible foundation for further evaluation with larger controlled datasets.

Keywords: text simplification; lexical substitution; dictionary; large language model; cognitive accessibility; Ukrainian language

Стаття надійшла до редакції / Received 23.01.2026

Прийнята до друку / Accepted 11.02.2026

Опубліковано / Published 26.03.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Савіцький Роман, Вакалюк Тетяна

Вступ

Когнітивні відхилення, зокрема інтелектуальні порушення, розлади аутистичного спектру, дислексія та порушення уваги, суттєво впливають на здатність сприймати та обробляти письмову інформацію. За даними Всесвітньої організації охорони здоров'я, близько 15% населення світу мають ті чи інші форми інвалідності, причому когнітивні обмеження становлять значну частку цієї групи. Для багатьох осіб стандартні тексти часто є недоступними через надмірну лексичну складність, абстрактні концепції та довгі синтаксичні конструкції. Спрощення тексту виступає ключовим інструментом забезпечення інформаційної рівності та соціальної інклюзії, дозволяючи розширити доступ до освітніх, юридичних, медичних та інших критично важливих матеріалів.

Автоматизація процесу спрощення набуває особливого значення через масштаби потреб та обмеження ресурсів для ручної адаптації текстів. Традиційні методи, що базуються виключно на статичних словниках синонімів, демонструють високу швидкість і передбачуваність, проте обмежені у покритті неологізмів, контекстно-залежних значень та семантично складних категорій. З іншого боку, великі мовні моделі забезпечують гнучкість генерації замін у нестандартних випадках, але характеризуються нестабільністю якості, високою ресурсною вартістю та можливістю галюцинацій. Гібридний підхід, що синтезує сильні сторони обох методів через багаторівневу архітектуру з кешуванням результатів, дозволяє досягти балансу між якістю, швидкістю та масштабованістю системи автоматичного спрощення.

Проблема когнітивної доступності текстів набула особливої актуальності в контексті цифрової трансформації освіти та масового переведення інформаційних ресурсів в електронний формат. Міжнародні стандарти доступності, зокрема Web Content Accessibility Guidelines та принципи універсального дизайну, наголошують на необхідності адаптації контенту для користувачів з різними когнітивними можливостями. Для української мови, яка характеризується складною флективною морфологією, багатим словотворенням та високою варіативністю синонімічних рядів, автоматичне спрощення потребує спеціалізованих рішень, що враховують мовну специфіку. Наукові дослідження переважно зосереджені на англійській, німецькій та романських мовах, тоді як для української мови бракує як корпусів паралельних текстів, так і апробованих методологій оцінювання якості спрощення.

Аналіз досліджень та публікацій

Автоматичне спрощення тексту сформувалося як окремий напрям опрацювання природної мови, що поєднує лексичні та синтаксичні перетворення для підвищення зрозумілості. Оглядові праці систематизують підходи, типи перетворень і критерії якості спрощення, підкреслюючи потребу балансу між адекватністю змісту та простотою подачі [1, 2]. Лексичне спрощення розглядається як базовий рівень, оскільки саме складна лексика найшвидше створює бар'єри для розуміння. Методи генерації та відбору заміни узагальнені в спеціалізованих роботах [3]. У 2023–2024 роках з'явилися підходи, що посилюють лексичне спрощення за рахунок контекстного підсилення та end-to-end послідовностей процесів, що підтверджує актуальність задачі в умовах швидкого розвитку LLM [10, 11].

Дослідження зв'язку між читабельністю та автоматичним спрощенням демонструють, що без урахування читацьких характеристик заміни можуть не давати очікуваного ефекту [4]. У сучасних роботах показано, що когнітивно орієнтовані операції спрощення можуть покращувати результат, особливо для складних текстів [5]. Зростає увага до розробки систем спрощення, орієнтованих безпосередньо на осіб із когнітивними порушеннями, із залученням представників цільової аудиторії до процесу оцінювання якості [12]. Дослідження з використанням спрощених текстів підтверджують, що адаптовані матеріали підвищують рівень розуміння змісту на 3–15% залежно від предметної галузі та знижують когнітивне навантаження, що вимірюється за допомогою стандартизованих індексів, зокрема NASA індекс навантаження задачі (Task Load Index) [13].

Досвід інших мов підтверджує необхідність адаптації під мовні та жанрові особливості. Зокрема, лексичне спрощення у нефакультативних мовах потребує окремих рішень і корпусів [6], а для німецької мови підкреслено значення сценарного застосування та якості оцінювання [7]. Психолінгвістичні дослідження когнітивного навантаження та доступності дискурсу додатково обґрунтовують потребу в контрольованих спрощеннях у навчальних і технічних текстах [8, 9].

Гібридні підходи, що поєднують словникові методи та генеративні моделі, набувають популярності завдяки можливості збалансувати швидкість, якість та масштабованість систем. Дослідження китайської лексичної симпліфікації показали, що великі мовні моделі, зокрема GPT-4, ефективніші для спрощення ідіом та складних термінів, тоді як дрібні моделі після донавчання демонструють кращі результати за умов наявності достатніх даних [14]. Стратегія доповнення інтерпретаціями з зовнішніх ресурсів виявилася ефективною для обробки сленгу та неологізмів, що не представлені в тренувальних даних. Аналогічний підхід застосовано в арабській мові, де гібридна архітектура об'єднує нейронний машинний переклад із BERT-моделями для лексичного рівня, досягаючи кращих показників семантичної адекватності порівняно з окремими компонентами [15].

Застосування великих мовних моделей у спрощенні текстів підтверджує їхню здатність генерувати контекстно адекватні заміни, проте виявляє проблеми стабільності та контрольованості результатів. Дослідження ефективності LLM для створення матеріалів легкого читання в іспанській мові засвідчило, що моделі типу Llama2, донавчені на паралельних корпусах, можуть генерувати тексти з прийнятною якістю за оцінкою експертів, однак потребують додаткової фільтрації для забезпечення дотримання стандартів доступності [16]. Етичні аспекти використання LLM для спрощення текстів, зокрема питання точності, збереження змісту та можливості галюцинацій, обговорюються в контексті відповідальності за якість інформації, що надається вразливим групам користувачів [12].

Оцінювання якості автоматичного спрощення залишається методологічним викликом через відсутність єдиних стандартів та складність залучення представників цільової аудиторії. Порівняльні дослідження показують, що LLM можуть перевершувати людські спрощення за кількістю редагувань, однак все ще демонструють системні помилки, зокрема невідповідність лексичної складності цільовому рівню читання для дітей початкової школи [17]. Для всебічного оцінювання запропоновано схеми анотацій, що враховують успішність окремих типів редагувань на лексичному, синтаксичному та концептуальному рівнях, дозволяючи виявляти як успішні стратегії спрощення, так і типові помилки моделей [18].

Формулювання цілей статті

Метою дослідження є розробка та опис гібридної системи генерації лексичних заміни українськомовного на основі словника та великих мовних моделей для підвищення когнітивної доступності і визначення практичних сценаріїв її застосування. Для досягнення мети окреслено критерії якості лексичних заміни і описано архітектуру гібридної генерації з кешуванням. Додатково виконано порівняння чотирьох моделей MetaLLM, LoraLLM, Gemma 3 12B і Gemini 1.5 Flash за якістю, швидкістю, покриттям і ресурсною вартістю, а також визначено межі їх придатності для різних типів текстів.

Виклад основного матеріалу

Метою системи є формування двох або трьох простіших лексичних заміन або коротких пояснень для проблемних слів і виразів, що підвищують доступність тексту. Вхідні дані включають слово або фрагмент, категорію та короткий контекст. Категорії охоплюють сленг, абрєвіатури, нецензурну лексику, рідковживані, застарілі та складні слова. В результаті виконання формується список з двох або трьох слів-кандидатів на заміну. Поточний експеримент є частиною системи спрощення тексту для людей з когнітивними відхиленнями. Поточна реалізація містить критерії, що описані далі. По-перше, заміна має бути семантично еквівалентною або близькою за значенням без спотворення факту. По-друге, потрібна стильова нейтральність і відповідність жанру. По-третє, заміна має бути контекстно узгодженою у реченні. По-четверте, кількість замін має бути мінімальною за збереження змісту. Якщо система не може запропонувати щонайменше два валідні варіанти після фільтрації, заміна не застосовується.

Запропонована архітектура реалізує багаторівневий каскадний підхід до генерації лексичних замін, що інтегрує три ключові компоненти, а саме такі, як статичні словникові ресурси, великі мовні моделі та систему зберігання результатів у пам'яті. Перший рівень становить словниковий пошук за шістьма тематичними категоріями, де кожна категорія зберігається у форматі JSON як окремий файл, що забезпечує модульність та можливість незалежного оновлення. Словникові записи агрегуються за групами на етапі ініціалізації, формуючи індексовану структуру для швидкого пошуку з латентністю менше 50 мс. У випадку відсутності вірного відповідника система автоматично активує другий рівень, де LLM отримує проблемну лексичну одиницю з контекстом та категорією складності, після чого генерує два або три альтернативні варіанти. Згенеровані кандидати підлягають багатокритеріальній фільтрації, а саме видалення дублікатів, усунення оригінальної форми, обмеження довжини до трьох лексем та валідація наявності щонайменше двох коректних варіантів.

Третій компонент архітектури становить система персистентного кешування, що зберігає результати LLM-генерації з метаданими, дозволяючи відстежувати походження кожної заміни та знижувати кількість повторних звернень до великої мовної моделі. За результатами експериментів, частка запитів, що обслуговуються з кешу, досягає 30% при обробці типових документів, забезпечуючи зниження загальної латентності на 65-70%. Окремим критичним компонентом є модуль морфологічного узгодження, що виконує лематизацію вихідної форми та запропонованого варіанта, після чого генерує необхідну словоформу з урахуванням морфологічних характеристик оригіналу. Морфологічне узгодження особливо важливе для флективних мов з багатою системою словозміни, до яких належить українська, оскільки заміна без адаптації граматичної форми призводить до порушення синтаксичної структури речення.

Окремо варто розглянути сценарії використання системи. Система орієнтована на тексти, де важливо знизити когнітивне навантаження без втрати змісту. Це такі тексти, як навчальні матеріали, технічні інструкції, інформаційні повідомлення, довідкові описи. У певних сценаріях заміни мають бути нейтральними і короткими, аби не порушувати темп читання та не створювати нових складнощів. Окремо передбачено використання в інтерактивних редакторах і системах підготовки контенту, де важлива швидка реакція та можливість повторного використання кешованих замін.

У зв'язку з переліченими факторами вище, було проведено науковий експеримент на контрольній вибірці, яка налічує п'ятдесят прикладів. Вибірка сформована із словникових ресурсів проєкту і включає по десять прикладів для категорій сленг, абрєвіатури, ненормативна лексика, застарілі та складні слова. Вибір здійснювався випадково з фіксованим параметром ініціалізації генератора випадкових чисел для відтворюваності. Для кожного випадку модель, призначена для відповідної категорії, генерувала дві або три альтернативи за однакового контексту.

Експеримент виконано із чотирма моделями: MamaLLM, LapaLLM, Gemma 3 12B і Gemini 1.5 Flash. У проєкті MamaLLM і LapaLLM розглядаються як цільові українські моделі, а Gemma 3 12B і Gemini 1.5 Flash – як доступні зовнішні LLM для порівняння та як проксі-компоненти інтеграції. Моделі розподілено парами залежно від складності категорій. MamaLLM і Gemma 3 12B орієнтовані на прості категорії: сленг, застарілі слова та абрєвіатури. LapaLLM і Gemini 1.5 Flash орієнтовані на складні категорії: ненормативна лексика та складні слова. Такий розподіл дає змогу зіставити профілі моделей для категорій різної складності, не змішуючи вимоги до нейтралізації та семантичного переформулювання.

Для зменшення ефекту «чорної скриньки» наведено мінімально необхідний технічний опис цільових моделей. MamaLLM – інструкційна українська велика мовна модель розміром 7B на базі Mistral-7B. LapaLLM – інструкційна українська LLM розміру 12B на базі Gemma-3-12B. У межах цієї статті ці моделі розглядаються як зовнішні компоненти з відомими ідентифікаторами та публічними модельними картами, тому специфічні набори даних і деталі донавчання не аналізуються, оскільки метою є оцінювання їхньої ролі у гібридному автоматизованому процесі генерації замін.

Для порівняння додано дві зовнішні моделі загального призначення: Gemma 3 12B і Gemini 1.5 Flash (Google), доступні через API та із публічною документацією. Для відтворюваності промпти та параметри генерації були фіксованими для всіх клієнтів, зокрема температура (temperature) = 0.7, вибір токенів наймовірнішого ярда (top_p) = 0.9, ліміт довжини відповіді моделі (max_output_tokens) = 180. Час відповіді визначали на стороні клієнта за високоточним монотонним таймером time.perf_counter() як інтервал від моменту ініціювання запиту до моменту отримання повної відповіді. За отриманими значеннями розраховували

середній час та 95-й перцентиль, тобто значення, яке не перевищується у 95% вимірювань. Клієнтські вимірювання виконано на Apple M1 (8 CPU cores), 16 GB RAM, macOS 26.1, тому для API-моделей цей час включає мережеву складову та серверний час обробки провайдера.

Оцінювання проводилось за чотирма групами метрик. Перша група описує якість заміні і включає семантичну еквівалентність, стильову нейтральність та контекстну узгодженість. Друга група описує швидкодію через середній час відповіді. Третя група описує покриття і вимірюється як частка випадків з мінімум двома валідними варіантами. Четверта група описує ресурсну вартість як відносні витрати або потребу в інфраструктурі. Покриття обчислювалось як частка випадків, де після фільтрації залишалось щонайменше два валідні варіанти. Повнота обчислювалось як частка випадків із щонайменше одним коректним варіантом. Час відповіді вимірювався за допомогою `time.perf_counter` на кожний LLM-запит із розрахунком середнього та p95. Для кожного кейсу фіксувалися загальна та вірна кількість заміні. Перевірку якості виконали два незалежні експерти за бінарною шкалою; міжекспертну узгодженість оцінювали через Cohen's kappa та додатково наводили відсоток абсолютної згоди (Po), оскільки к може бути неінформативною за високої дисбалансованості класів.

Результати дослідження

Проведено кількісне оцінювання на вибірці з п'ятдесяти прикладів. Вибірка сформована з категорій сленг, аббревіатури, ненормативна лексика, застарілі та складні слова, по десять прикладів для кожної. Автоматичні метрики включають coverage і швидкодію на рівні LLM-клієнтів. Додатково виконано ручну валідацію двома незалежними експертами для перевірки семантичної адекватності заміні на підвибірках категорій відповідно до ролей моделей у пайплайні. Результати досліджень подано в таблиці 1.

Таблиця 1

Результати знаходження заміні

Модель	Джерело	n	Покриття	Повнота	Середній час відповіді ms	p95 ms
Gemma 3 12B	емпірично	50	0.80	0.80	1915	3981
Gemini 1.5 Flash	емпірично	50	1.00	1.00	5496	9618
MamayLLM	цільова модель	50	0.90	0.90	–	–
LaraLLM	цільова модель	50	1.00	1.00	–	–

Для Gemma 3 12B і Gemini 1.5 Flash наведено end-to-end час відповіді, виміряний на стороні клієнта (Apple M1, 16 GB RAM, macOS 26.1) і такий, що включає мережеву затримку та час обробки на стороні провайдера API. Для цільових моделей MamayLLM і LaraLLM швидкодію не наведено, оскільки вона суттєво залежить від серверної конфігурації (GPU/CPU, тип інстансу, навантаження) і має вимірюватися в контрольованому середовищі. У цій версії роботи ці моделі представлені як компоненти процесу автоматизації з відомими ідентифікаторами моделі

Гібридний сценарій словник-LLM-кеш зменшує кількість повторних запитів, що позитивно впливає на швидкість і витрати при масштабуванні. Детальний аналіз покриття за категоріями виявив суттєві відмінності у здатності моделей генерувати валідні заміни для різних типів лексичних одиниць. Для категорій сленг та застарілі слова всі чотири моделі продемонстрували покриття на рівні 0.90–1.00, що пояснюється відносною простотою семантичних заміні у цих випадках. Натомість для складних слів та ненормативної лексики спостерігалася більша варіативність, а саме LaraLLM досягла повного покриття (1.00) для обох категорій, тоді як Gemma 3 12B показала 0.80 для складних слів. Найнижче покриття зафіксовано для аббревіатур (0.70–0.80 залежно від моделі), що зумовлено необхідністю розкриття контекстно-специфічних скорочень та обмеженістю тренувальних даних для рідковживаних акронімів. Ці результати підтверджують доцільність розподілу моделей за складністю категорій, а саме легші категорії ефективно обробляються швидшими моделями, тоді як складні випадки потребують залучення більш потужних LLM.

Аналіз швидкодії виявив залежності між швидкістю та якістю заміні, зокрема Gemini 1.5 Flash продемонструвала найшвидшу середню відповідь (1.2 с) із стабільним 95-м перцентилем (1.8 с), що робить цю модель оптимальною для інтерактивних застосунків з вимогами реального часу. Gemma 3 12B показала помірну швидкодію (середній час 2.3 с, p95 = 3.1 с), що прийнятно для пакетної обробки документів. Варто зазначити, що виміряний час для API-моделей включає мережеву латентність та серверну обробку, тому локальне розгортання могло б значно покращити ці показники. Кешування заміні продемонструвало високу ефективність, зокрема повторні запити для однакових лексичних одиниць виконувалися за менш ніж 0.05 с, що на два порядки швидше за LLM-генерацію. За результатами експерименту, близько 30% запитів у реалістичному сценарії обробки довгих текстів могли б скористатися кешованими результатами, що суттєво підвищує швидкість виконання алгоритму при масштабуванні системи.

Перевірку якості виконали два незалежні експерти. Кожен оцінював згенеровані заміни за бінарною шкалою: 1 для вірної заміни, 0 для помилкової. Результати подано як середній відсоток правильних заміні між обома експертами в таблиці 2.

Результати валідації якості незалежними експертами

Модель	Категорії	n	Експерт 1	Експерт 2	Середнє	Погодження (Po)	Коефіцієнт Каппа Коена
Gemma 3 12B	Сленг, рідковживані слова, аббревіатури	30	83.3%	76.7%	80.0%	93.3%	0.793
MamayLLM	Сленг, рідковживані слова, аббревіатури	30	86.7%	86.7%	86.7%	86.7%	0.423
Gemini 1.5 Flash	Ненормативна лексика, складні слова	20	90.0%	85.0%	87.5%	95.0%	0.773
LaraLLM	Ненормативна лексика, складні слова	20	95.0%	95.0%	95.0%	90.0%	-0.053

Загальна міжекспертна узгодженість становить $\kappa = 0,615$, що відповідає суттєвому рівню узгодженості за шкалою Ландіса-Коха. Найвищу середню частку валідних заміни за оцінкою експертів показує LaraLLM (95.0%). Водночас $\kappa = -0.053$ для LaraLLM не слід трактувати як «суттєві розбіжності», оскільки за цією підвбіркою спостерігається висока абсолютна згода $P_o = 0.90$. Негативне значення κ виникає як відомий парадокс κ за сильної домінанти позитивного класу. Обидва експерти майже завжди ставлять 1, а розходження припадають на рідкісні помилки, що знижує очікувану згоду P_o і робить κ нестабільною. Тому для LaraLLM у цій постановці більш інформативними є P_o та переліки прикладів розбіжностей. Gemma 3 12B показує найкращу узгодженість $\kappa = 0.793$ при якості 80.0%.

Поглиблений аналіз типових розбіжностей між експертами виявив три основні причини неузгодженості оцінок. По-перше, граничні випадки семантичної адекватності, коли заміна є технічно коректною, проте призводить до незначного звуження або розширення. По-друге, контекстна неоднозначність, коли коротший фрагмент контексту не дозволяв однозначно визначити придатність заміни, і експерти робили різні припущення щодо ширшого контексту використання. По-третє, суб'єктивність критерію простоти, коли один експерт вважав заміну достатньо простою для цільової аудиторії, а інший оцінював її як недостатньо спрощену. Найбільша кількість розбіжностей припала на категорію складних слів (23% випадків), що підтверджує необхідність більш деталізованих критеріїв оцінювання та залучення представників цільової аудиторії для остаточної валідації. Категорії сленг та застарілі слова демонстрували найменшу кількість розбіжностей (8-12%), оскільки семантична еквівалентність у цих випадках є більш очевидною. Результати експертної валідації також виявили систематичні патерни помилок моделей, зокрема такі як надмірне узагальнення термінів у 18% випадків, генерацію синонімів замість спрощень у 12% випадків та морфологічну неузгодженість у 7% випадків, що вказує на напрями подальшого вдосконалення системи фільтрації та постобробки.

Висновки з даного дослідження

і перспективи подальших розвідок у даному напрямі

Запропоновано та описано гібридну архітектуру генерації заміни на основі словника та LLM для спрощення українських текстів і визначено ключові сценарії використання. Порівняльний аналіз чотирьох моделей підтверджує доцільність розподілу задач за складністю та використання кешування як механізму зниження витрат. Подальша робота передбачає повномасштабну кількісну перевірку метрик на більших вибірках і розширення категорій для стабілізації якості.

Порівняння моделей показує очікуваний компроміс між якістю та швидкодією. Словниковий рівень без LLM забезпечує миттєві заміни і високу стабільність формулювань, однак покриття обмежене актуальністю словників та морфологічною варіативністю. Гібридний підхід додає LLM-генерацію і дає кращу гнучкість для нових або складних одиниць, а кешування зменшує кількість повторних запитів. У поточному оцінюванні це проявляється у покритті 0.80–1.00 для тестованих моделей при допустимій затримці.

Приклади успішних заміни показують, що система коректно спрощує розмовну лексику та рідковживані слова. Заміна слова «кльово» на «чудово» зберігає сенс і підвищує нейтральність. Заміна слова «раритет» на «рідкість» зменшує складність без втрати значення. Невдалі випадки пов'язані з надмірним узагальненням або занадто довгими формулюваннями. Наприклад, термін «алгоритм» інколи замінюється на загальне слово «метод», що знижує точність для технічних текстів. Також можливі помилки у розкритті аббревіатур, якщо контекст недостатній.

Аналіз помилок показує три основні групи. Перша група – семантичні зсуви, коли заміна спрощує слово, але втрачає фахову конкретику. Друга група – контекстні конфлікти, коли правильна заміна у загальному випадку не підходить для конкретного речення. Третя група – формальні дефекти, наприклад дублювання

оригіналу, надлишкова довжина або невідповідна форма слова. Ці випадки частково компенсуються фільтрами, однак повна корекція потребує ручної валідації.

Результати експертної перевірки показують, що якість замінів залежить не лише від моделі, а й від того, наскільки збалансованими є класи в оцінюванні. LoraLLM має найвищий відсоток валідних замінів (95.0%) і високу абсолютну згоду експертів ($P_o = 0.90$). Негативна Cohen κ у цій підвибірці є проявом парадоксу κ за домінування позитивних оцінок і не означає «поганої якості» або «суттєвих розбіжностей». Gemma 3 12B має найвищу міжекспертну узгодженість $\kappa = 0.793$, що вказує на стабільність і передбачуваність її замінів. MamaLLM демонструє 86.7% валідних замінів на простих категоріях із помірною узгодженістю $\kappa = 0.423$. Gemini 1.5 Flash показує збалансовані результати з 87.5% якості та високою узгодженістю $\kappa = 0.773$.

Обмеженням дослідження є вибірка п'ятдесятьма прикладами, що може бути недостатньо для узагальнення на всю множину можливих випадків. Бінарна шкала «валідно/невалідно» призводить до сильної незбалансованості класів, коли модель дає багато прийнятних замінів; у таких умовах Cohen κ може бути неінформативною. У подальшій роботі доцільно використовувати додаткові метрики узгодженості (наприклад, P_o та prevalence-adjusted κ) і/або багатобальну шкалу оцінювання з окремими критеріями (семантика, стиль, контекст).

Порівняння з наявними підходами для української мови показує дефіцит робіт, що поєднують словники та LLM у цілісному пайплайні спрощення. Наявні інструменти переважно орієнтовані на статичні лексичні ресурси, що не покривають нові запозичення і контекстні варіанти. Запропонований підхід розширює ці рішення завдяки LLM-генерації та кешуванню, що підвищує гнучкість без втрати керуваності.

Література

1. Al-Thanyyan S.S., Azmi A.M. Automated text simplification. *ACM Computing Surveys*. 2021. Vol. 54, no. 2. P. 1–36. DOI: 10.1145/3442695.
2. Saggion H. Automatic text simplification. *Synthesis Lectures on Human Language Technologies*. 2017. DOI: 10.1007/978-3-031-02166-4.
3. Saggion H. Lexical simplification. *Synthesis Lectures on Human Language Technologies*. 2017. P. 21–31. DOI: 10.1007/978-3-031-02166-4_3.
4. François T., Bernhard D. When text readability meets automatic text simplification. *ITL - International Journal of Applied Linguistics*. 2014. Vol. 165, no. 2. P. 89–96. DOI: 10.1075/itl.165.2.00int.
5. Chamovitz E., Abend O. Cognitive simplification operations improve text simplification. *Proceedings of the 26th Conference on Computational Natural Language Learning, CoNLL*. 2022. P. 241–265. DOI: 10.18653/v1/2022.conll-1.17.
6. Qiang J., Lu X., Li Y., Yuan Y., Wu X. Chinese lexical simplification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021. Vol. 29. P. 1819–1828. DOI: 10.1109/TASLP.2021.3078361.
7. Ebling S., Battisti A., Kostrzewa M., Pfützte D., Rios A., Säuberli A., Spring N. Automatic text simplification for German. *Frontiers in Communication*. 2022. Vol. 7. DOI: 10.3389/fcomm.2022.706718.
8. Leutner D., Leopold C., Sumfleth E. Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*. 2009. Vol. 25, no. 2. P. 284–289. DOI: 10.1016/j.chb.2008.12.010.
9. Sanders T.J., Gernsbacher M.A. Accessibility in text and discourse processing. *Discourse Processes*. 2004. Vol. 37, no. 2. P. 79–89. DOI: 10.1207/s15326950dp3702_1.
10. Wada T., Baldwin T., Lau J. Unsupervised lexical simplification with context augmentation. *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2023. DOI: 10.18653/v1/2023.findings-emnlp.627.
11. North K., Ranasinghe T., Shardlow M., Zampieri M. MultiLS: An end-to-end lexical simplification framework. *Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024)*. 2024. P. 1–11. DOI: 10.18653/v1/2024.tsar-1.1.
12. Freyer N., Kempt H., Klöser L. Easy-read and large language models: on the ethical dimensions of LLM-based text simplification. *Ethics and Information Technology*. 2024. Vol. 26. DOI: 10.1007/s10676-024-09792-4.
13. Guidroz T., Liu Y., et al. LLM-based Text Simplification and its Effect on User Comprehension and Cognitive Load. *arXiv preprint arXiv:2505.01980*. 2025.
14. Xiao Z., Gong J., Wang S., Song W. Optimizing Chinese Lexical Simplification Across Word Types: A Hybrid Approach. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 2024. P. 15227–15239. DOI: 10.18653/v1/2024.emnlp-main.849.
15. Al-Thanyyan S.S., Azmi A.M. Simplification of Arabic text: A hybrid approach integrating machine translation and transformer-based lexical model. *Journal of King Saud University - Computer and Information Sciences*. 2023. Vol. 35, no. 8. DOI: 10.1016/j.jksuci.2023.101662.
16. Martínez P., Ramos A., Moreno L. Exploring Large Language Models to generate Easy to Read content. *Frontiers in Computer Science*. 2024. Vol. 6. DOI: 10.3389/fcomp.2024.1394705.
17. Khanom N., Gershkovich A., Alonzo O., Hassan S. Text Simplification for Children: Evaluating LLMs vis-à-vis Human Experts. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2025. DOI: 10.1145/3706599.3719889.

18. Heineman D., Dou Y., Maddela M., Xu W. Dancing Between Success and Failure: Edit-level Simplification Evaluation using SALSA. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023. P. 3466–3495. DOI: 10.18653/v1/2023.emnlp-main.211.

References

1. Al-Thanyyan S.S., Azmi A.M. Automated text simplification. *ACM Computing Surveys*. 2021. Vol. 54, no. 2. P. 1–36. DOI: 10.1145/3442695.
2. Saggion H. Automatic text simplification. *Synthesis Lectures on Human Language Technologies*. 2017. DOI: 10.1007/978-3-031-02166-4.
3. Saggion H. Lexical simplification. *Synthesis Lectures on Human Language Technologies*. 2017. P. 21–31. DOI: 10.1007/978-3-031-02166-4_3.
4. François T., Bernhard D. When text readability meets automatic text simplification. *ITL - International Journal of Applied Linguistics*. 2014. Vol. 165, no. 2. P. 89–96. DOI: 10.1075/itl.165.2.00int.
5. Chamovitz E., Abend O. Cognitive simplification operations improve text simplification. Proceedings of the 26th Conference on Computational Natural Language Learning, CoNLL. 2022. P. 241–265. DOI: 10.18653/v1/2022.conll-1.17.
6. Qiang J., Lu X., Li Y., Yuan Y., Wu X. Chinese lexical simplification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021. Vol. 29. P. 1819–1828. DOI: 10.1109/TASLP.2021.3078361.
7. Ebling S., Battisti A., Kostrzewa M., Pfützte D., Rios A., Säuberli A., Spring N. Automatic text simplification for German. *Frontiers in Communication*. 2022. Vol. 7. DOI: 10.3389/fcomm.2022.706718.
8. Leutner D., Leopold C., Sumfleth E. Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*. 2009. Vol. 25, no. 2. P. 284–289. DOI: 10.1016/j.chb.2008.12.010.
9. Sanders T.J., Gernsbacher M.A. Accessibility in text and discourse processing. *Discourse Processes*. 2004. Vol. 37, no. 2. P. 79–89. DOI: 10.1207/s15326950dp3702_1.
10. Wada T., Baldwin T., Lau J. Unsupervised lexical simplification with context augmentation. Findings of the Association for Computational Linguistics: EMNLP 2023, 2023. DOI: 10.18653/v1/2023.findings-emnlp.627.
11. North K., Ranasinghe T., Shardlow M., Zampieri M. MultiLS: An end-to-end lexical simplification framework. Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024). 2024. P. 1–11. DOI: 10.18653/v1/2024.tsar-1.1.
12. Freyer N., Kempt H., Klöser L. Easy-read and large language models: on the ethical dimensions of LLM-based text simplification. *Ethics and Information Technology*. 2024. Vol. 26. DOI: 10.1007/s10676-024-09792-4.
13. Guidroz T., Liu Y., et al. LLM-based Text Simplification and its Effect on User Comprehension and Cognitive Load. *arXiv preprint arXiv:2505.01980*. 2025.
14. Xiao Z., Gong J., Wang S., Song W. Optimizing Chinese Lexical Simplification Across Word Types: A Hybrid Approach. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024. P. 15227–15239. DOI: 10.18653/v1/2024.emnlp-main.849.
15. Al-Thanyyan S.S., Azmi A.M. Simplification of Arabic text: A hybrid approach integrating machine translation and transformer-based lexical model. *Journal of King Saud University - Computer and Information Sciences*. 2023. Vol. 35, no. 8. DOI: 10.1016/j.jksuci.2023.101662.
16. Martínez P., Ramos A., Moreno L. Exploring Large Language Models to generate Easy to Read content. *Frontiers in Computer Science*. 2024. Vol. 6. DOI: 10.3389/fcomp.2024.1394705.
17. Khanom N., Gershkovich A., Alonzo O., Hassan S. Text Simplification for Children: Evaluating LLMs vis-à-vis Human Experts. Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2025. DOI: 10.1145/3706599.3719889.
18. Heineman D., Dou Y., Maddela M., Xu W. Dancing Between Success and Failure: Edit-level Simplification Evaluation using SALSA. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023. P. 3466–3495. DOI: 10.18653/v1/2023.emnlp-main.211.