

Іваницька Н. Б.<https://orcid.org/0000-0002-9925-1285>

Вінницький торговельно-економічний інститут Державного торговельно-економічного університету

КОРПУСНИЙ АНАЛІЗ МОВНОЇ АСИМЕТРІЇ КОНСТРУКЦІЙ ІЗ ДЕСЕМАНТИЗОВАНИМИ ДІЄСЛОВАМИ У GPT-ПЕРЕКЛАДІ МЕДІАТЕКСТІВ

У статті подано результати корпусного аналізу мовної асиметрії десемантизованих дієслів у перекладі медійних текстів великими мовними моделями (ВММ), зокрема чатом GPT. Основну увагу зосереджено на специфіці функціонування та відтворення англійських конструкцій із десемантизованими дієслівними компонентами (КДД), зокрема таких продуктивних одиниць, як *do, take, take, give, have, get* у контексті англо-українського машинного перекладу. Матеріалом для дослідження слугував спеціально сформований паралельний корпус, що містив оригінальні публікації жанру *Lifestyle* британського видання «*The Guardian*» та їхній переклад, генерований ВММ (чатом GPT). Актуальність наукового пошуку зумовлена виразною структурно-семантичною асиметрією між аналітичною природою англійських КДД, де лексичне значення сконцентроване в іменному компоненті, а дієслово виконує переважно граматичну функцію, та синтетичним ладом української мови, яка тяжіє до використання повнозначних дієслівних предикатів. Зазначена міжмовна невідповідність створює особливі когнітивні та алгоритмічні виклики для генеративних систем перекладу, що потребує детального вивчення здатності ВММ адаптувати ці конструкції до норм цільової мови. Методологічний апарат дослідження базувався на використанні професійного інструментарію *Sketch Engine*, який уможливив точне вирівнювання сегментів паралельного корпусу, автоматизовану екстракцію релевантних конструкцій та їх подальшу класифікацію за перекладацькими стратегіями. Наукова вибірка охопила 127 верифікованих контекстів. Результати корпусного аналізу засвідчили, що домінантною стратегією подолання асиметрії виступає синтетична вербалізація (45% випадків). Це демонструє високий рівень здатності ВММ інтегрувати десемантизоване дієслово та іменник у єдиний синтетичний предикат української мови, забезпечуючи прагматичну точність і стилістичну відповідність медіатексту. Поряд із цим, у роботі детально описано застосування стратегій функціональної заміни, перифразу (експлікації), а також зафіксовано випадки буквального покомпонентного перекладу, які становлять зону ризику для якості машинного перекладу. Отримані дані підтверджують гіпотезу про прогресуючу адаптивність ВММ до складних випадків міжмовної асиметрії в межах медійного дискурсу.

Ключові слова: десемантизовані дієслова, англо-український генеративний машинний переклад, великий мовний модедь, паралельний корпус, медійний дискурс, перекладацькі стратегії, корпусний аналіз, аналітичні конструкції, міжмовна асиметрія.

Постановка проблеми. У сучасній англомовній лінгвістичній традиції номінацію «*light verbs*» (інші терміни – «*delexical verb*», «*semantically weak verb*», «*empty verb*», «*stretched verb*») використовують для позначення дієслів із загальним, широким значенням, яке актуалізується або уточнюється в поєднанні з іншим словом. Як зауважує Дж. Ліч, «*light verb* – поширене й універсальне лексичне дієслово (*do, give, have, take* або *take*), яке у багатьох своїх вживаннях є семантично

«слабким» і може поєднуватися з іменниками в конструкціях на кшталт *do the cleaning, give (someone) a hug, have a drink, make a decision, take a break* [13, с. 254]. Такі конструкції традиційно кваліфікують як «*light verb constructions*» (LVC) [18; 19].

В україністиці явища, подібні до LVC, розглядають крізь призму десемантизації дієслівного компонента. Зокрема, К. Городенська та І. Вихованець наголошують, що в таких одиницях дієслово



втрачає лексичну самостійність, виконуючи суто граматичну роль у межах аналітичної конструкції [2; 3]. А. Загнітко у своїх роботах використовує терміни «делексикалізація» та «десемантизація» [4, с. 310]. Згадані конструкції також кваліфікують аналітичними дієсловами, називаючи дієслівні компоненти лексемами широкої семантики [5; 6]. У межах цієї розвідки послуговуємося термінами «десемантизоване дієслово» (ДД) та «комплекс із десемантизованим дієсловом» (КДД) з огляду на об'єкт та мету дослідження.

Доведено, що ДД у перекладі є специфічними через структурно-семантичну асиметрію: англійські конструкції реалізують аналітичну модель предикації, де семантика зосереджена в іменному компоненті, тоді як українська мова надає перевагу синтетичній вербалізації. Це спричиняє перекладацькі зсуви, врахування яких не завжди відтворено алгоритмами генеративних систем машинного перекладу (МП), що, у свою чергу, призводить до вибору стилістично неприродних відповідників. Додамо також, що міжмовна асиметрія є критичною для англо-українського перекладу, оскільки українська мова зазвичай вимагає трансформації англійського аналітизму в компактний синтетичний відповідник (*take a look* → *глянути*). Можемо припустити, що здатність ВММ до такої функціональної заміни на противагу покомпонентному буквальному перекладу є одним із показників якості перекладу.

Аналіз останніх досліджень і публікацій показує, що конструкції з десемантизованими дієсловами мають глибоке осмислення в теоретичній граматиці, семантиці та корпусній лінгвістиці. В українській мовознавчій традиції їх називають дієслівно-іменниковими аналітичними предикатами і описують із позицій десемантизації та делексикалізації дієслівного компонента [3], аналітичних тенденцій у граматичній системі мови [2, с. 4] та функціонального перерозподілу семантичних ролей [1; 4]. Згадані напрацювання формують теоретичне підґрунтя для подальшого міжмовного аналізу КДД.

В англомовній лінгвістиці КДД докладно описані з позицій синтаксичної теорії [11; 16; 19], семантики [14; 15], типологічної варіативності [17; 19] та корпусних методів [12]. Ці дослідження виявляють структурну та семантичну складність КДД, їхню варіативність і високу частотність у сучасному англомовному дискурсі, однак зосереджуються переважно на описі мовної системи.

Перекладознавчий аспект аналітичних дієслів частково висвітлено у працях автора [5; 6] із

акцентом на міжмовній асиметрії між аналітичними конструкціями англійської мови та синтетичними дієслівними відповідниками української. Проте такі одиниці не розглядалися в аспекті системних трансформацій під час генеративного МП, як і не застосовувалися корпусні методи їхнього дослідження для фіксації та інтерпретації системних випадків міжмовної асиметрії та шляхів її подолання.

Отже, окремі аспекти міжмовної асиметрії аналітичних дієслів уже висвітлювалися в наукових працях, однак вони здебільшого стосувалися людського перекладу або систем МП попередніх поколінь. Натомість специфіка системних трансформацій десемантизованих конструкцій у межах генеративного МП (зокрема ВММ) залишається поза увагою дослідників. Відсутність комплексних корпусних даних, які б дозволили зафіксувати та інтерпретувати новітні стратегії подолання асиметрії алгоритмами ШІ, створює суттєву лакуну в сучасному перекладознавстві. Саме необхідність наукового осмислення того, як ВММ адаптують англійські аналітичні структури до синтетичного ладу української мови, зумовлює актуальність та новизну нашої роботи.

Постановка завдання. Метою роботи є комплексне виявлення та систематизація особливостей подолання мовної асиметрії під час генеративного машинного перекладу англійських конструкцій із десемантизованими дієсловами на українську мову. Об'єкт дослідження обмежений жанровими параметрами медійного дискурсу, а основним інструментом аналізу є корпусні технології.

Методологія дослідження. Методологічною основою дослідження є паралельний англо-український корпус медійних текстів розділу Lifestyle видання «The Guardian» та їхнього перекладу українською за допомогою ВММ (ChatGPT). Корпус охоплював 23 890 токенів (20 313 слів, 1 160 речень) в англійській частині та 17 295 токенів (13 683 слова, 1 048 речень) в українському перекладі, згенерованому чатом GPT. Корпусну розмітку, вирівнювання сегментів та екстракцію конструкцій здійснено за допомогою функціоналу *Sketch Engine*. Для подальшого аналізу було відібрано найбільш продуктивні десемантизовані дієслівні лексеми у корпусі: *do*, *make*, *take*, *give*, *have*, *get*.

Істотно, що методика дослідження базувалася на застосуванні мови корпусних запитів SQL у середовищі *Sketch Engine*, що дозволило цілеспрямовано екстрагувати англійські конструкції з десемантизованими дієсловами (*make*, *take*,

have, give, do, get) як первинні носії міжмовної асиметрії. Використання гнучких запитів із урахуванням дистантних компонентів забезпечило фіксацію аналітичних моделей оригіналу для їх подальшого зіставлення з перекладацькими відповідниками, що уможливило провести кількісну оцінку здатності ВММ долати структурну невідповідність через синтетизацію або функціональну заміну предикатів. Описаний підхід забезпечив високу валідність вибірки, що склала 127 верифікованих одиниць.

Виклад основного матеріалу. Аналіз отриманого фактичного матеріалу дозволив ідентифікувати низку перекладацьких трансформацій, які виступають ключовими інструментами адаптації та нівелювання мовної асиметрії в процесі відтворення англійських аналітичних структур українською мовою. Виявлені перекладацькі трансформації репрезентовані такими типами: 1) синтетична вербалізація (радикальне подолання асиметрії шляхом перетворення аналітичної конструкції вихідного тексту в повнозначне дієслово в українському перекладі), 2) функціональна заміна (нівелювання семантичної невідповідності через заміну ДД більш конкретним і типовим для мови перекладу предикатом), 3) перифраз / експлікація (розгортання структури для збереження прагматичного змісту в умовах міжмовної нееквівалентності), 4) буквальний покомпонентний переклад (випадок нездоланої асиметрії, за якого зберігається структура оригіналу без необхідної семантичної адаптації до норм української мови).

Далі наводимо отримані нами статистичні дані контекстуальних варіантів перекладу для визначення закономірностей у роботі ВММ та ділимося нашими інтерпретаційними спостереженнями.

Синтетична вербалізація є домінуючою стратегією перекладу в досліджуваному корпусі. За нашими підрахунками, в 58 випадках (45%) ВММ успішно розпізнає КДД як семантично цілісну одиницю та інтегрує ДД та іменник в синтетичну повнозначну дієслівну лексему, зберігаючи контекстуальні та прагматичні особливості вихідного тексту. Наведемо кілька прикладів досліджуваних нами конкордансів. У парі *On the morning of day four we got word that the snowplough was coming* – *Вранці четвертого дня ми дізналися, що приїде снігоприбиральна машина* ГРТ розпізнає «got word» як семантично цілісну дію «отримати інформацію». Іменник «word» виступає носієм акціонального центру, а десемантизоване дієслово «got» згортається в один повнозначний дієслівний предикат «дізналися», що

забезпечує синтетизацію та збереження прагматичної точності контексту. В іншому прикладі *She did the cleaning before guests arrived* – *Вона прибирала перед приходом гостей* конструкція із десемантизованим «did» конвертується в українську однослівну номінацію «прибирала».

У наступному прикладі *She took a morning walk along the river every day* – *Вона щодня прогулювалася вздовж річки вранці* ГРТ розпізнає акціональний центр іменника та трансформує його в дієслівну форму цільової мови. Крім того, алгоритм коректно передає аспект дії та частотний контекст («щодня»), що важливо для відтворення семантики регулярної активності в медійному жанрі. Отже, переклад видається не лише лексично адекватним, а й прагматично точним.

Функціональна заміна відбувається тоді, коли ВММ замінює десемантизоване дієслово більш конкретним, контекстуально зумовленим дієсловом, зберігаючи семантичний центр конструкції, зазвичай закладений у іменнику. Така трансформація дозволяє адаптувати переклад до норм української мови, не порушуючи стилістики та прагматики медійного тексту. У нашій вибірці було зафіксовано 35 випадків застосування такої трансформації (28% від усіх конкордансів). Проаналізуємо декілька прикладів. Так, у парі *She was invited to give a speech about the psychological benefits of mindfulness at the wellness conference* – *Її запросили виголосити промову про психологічні переваги усвідомленості на конференції з питань добробуту в англійській конструкції «give a speech» іменник «speech» виконує роль семантичного ядра, передаючи подію (сам виступ), тоді як десемантизоване «give» несе переважно граматичне навантаження. ВММ здійснює функціональну заміну give → виголосити, що перетворює конструкцію на функціонально еквівалентний двокомпонентний вираз. Семантичне ядро зберігається (speech → промова), а переклад стає природним для українського медійного стилю. У мовній парі *If you get a chance during the seminar, try to ask the speaker about their approach to digital wellbeing* – *Якщо випадє нагода під час семінару, спробуйте поставити доповідачу запитання щодо його підходу до цифрового добробуту* іменник «chance» є семантичним ядром (можливість здійснення дії), а десемантизоване «get» виражає граматичні категорії. ВММ здійснює функціональну заміну, адаптуючи вираз до норм української стилістики: «get → випадє».*

Не менш показовими виявилися випадки перифразу (експлікації), які охоплюють 17%

вибірці (21 фіксація). На відміну від прямої синтетичної вербалізації, де конструкція згортається до одного дієслова, перифраз у виконанні ВММ демонструє здатність алгоритмів до контекстуального розгортання семантики. Це не просто заміна лексем, а стратегія підвищення «ваги» висловлювання, яка нерідко виступає критично важливою для медійного дискурсу. ВММ може ідентифікувати англійську структуру як стилістично недостатню для українських новин і «добудувати» її, додаючи уточнювальні компоненти.

Проаналізуємо декілька зафіксованих у корпусі прикладів. Так, англійське речення *The international foundation gives help to local hospitals affected by the crisis* ВММ трансформує в українське *Міжнародний фонд надає необхідну підтримку місцевим лікарням, що постраждали від кризи*. У цьому разі використання перифразу «надає підтримку» замість нейтрального дієслова дозволяє підкреслити системність і офіційний характер допомоги, що повністю відповідає нормам української медійної комунікації. Аналогічна тенденція простежується в прикладі *The government must take action to curb rising inflation before the end of the quarter* – *Уряд має вжити рішучих заходів* задля стримування зростання інфляції до кінця кварталу. Замість функціонально точного, але стилістично спрощеного варіанта «має діяти», модель генерує перифраз: «має вжити рішучих заходів». Додавання означення «рішучих» до вихідного поняття «action» свідчить про те, що ВММ орієнтується на типові колокації українського медіа-поля, де боротьба з еконо-

мічними викликами традиційно описується через категорію «вжиття заходів». Такі результати підтверджують, що перифраз у перекладах ВММ є проявом прагматичної еквівалентності.

Стратегію калькування було зафіксовано в 10% випадків (12 прикладів) у досліджуваній вибірці. Виявлені випадки свідчать про певне технічне обмеження ВММ, за якого алгоритм не ідентифікує КДД як цілісну семантичну одиницю. Внаслідок цього десемантизоване дієслово помилково трактується як повнозначне, що призводить до порушення стилістичних норм української мови. У медійних текстах калькування створює штучні конструкції, які ускладнюють сприйняття інформації. Розглянемо найбільш показові випадки. Так, у паралельному конкордансі ... *I find this makes the situation even more unequal* – ...це *робить ситуацію* ще більш нерівною чат GPT здійснює прямий переклад «make → «робить», ігноруючи потребу в синтетичному або функціональному відповіднику (наприклад, «поглиблює нерівність»). Знаходимо й інший приклад із дослівним перекладом десемантизованого дієслова «have»: *Now Brooklyn says he and Peltz have a look at the property* – *Тепер Бруклін каже, що він та Пельтц мають погляд на нерухомість*. ВММ інтерпретує «have» як дієслово володіння, повністю втрачаючи ідіоматичну та прагматичну функцію КДД.

Узагальнені спостереження щодо релевантних та частотних стратегій подолання міжмовної асиметрії в досліджуваному дискурсі узагальнено на (рис. 1).

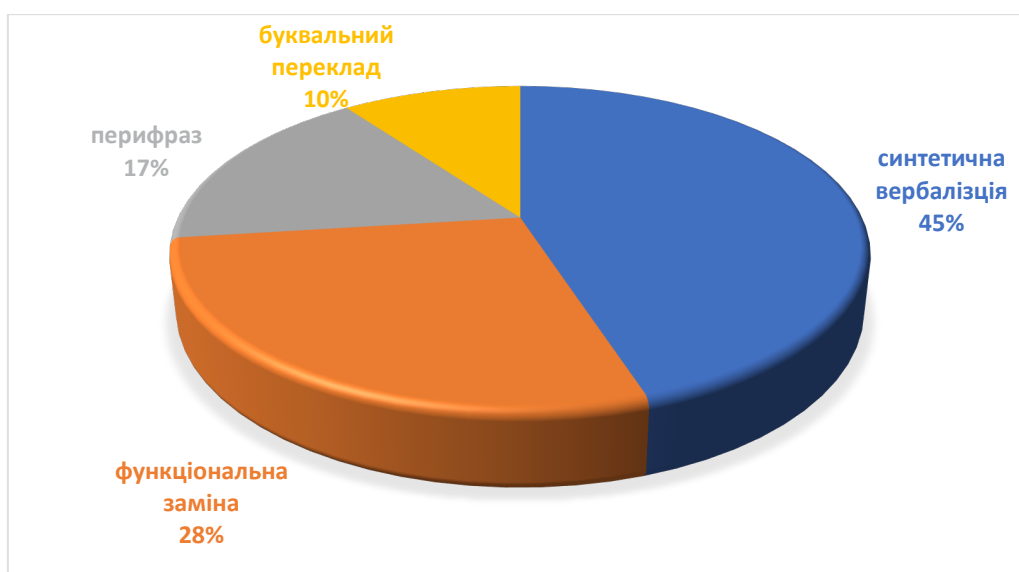


Рис. 1. Розподіл перекладацьких стратегій при відтворенні конструкцій із десемантизованими дієсловами в англо-українському медійному корпусі

Для отримання об'єктивної картини того, як ВММ долають структурну розбіжність між аналітизмом англійських КДД та синтетизмом української мови, було здійснено кількісний підрахунок застосованих трансформацій в розрізі кожного з десемантизованих дієслів (*do, make, take, give, have, get*). Узагальнені результати цього аналізу, що базуються на вибірці зі 127 одиниць, представлено в таблиці 1.

Результати кількісного аналізу у розрізі окремих десемантизованих дієслів дають підстави припустити, що лексична специфіка кожного дієслова безпосередньо впливає на вибір стратегії подолання мовної асиметрії. Найбільш репрезентативним у медійному корпусі виявилось дієслово *make* (36 одиниць), для якого характерна висока частотність синтетичної вербалізації (16 випадків), що пояснюється наявністю усталених дієслівних відповідників для таких конструкцій в українській мові. Водночас саме для *make* зафіксовано найбільшу кількість випадків буквального перекладу (4), що вказує на наявні проблемні алгоритми міжмовної адаптації. Дієслово *do* (30 одиниць) продемонструвало найвищий рівень успішного нівелювання асиметрії через сукупність синтетичної вербалізації та функціональної заміни (21 випадок). Стабільність показників для дієслова *take* (22 одиниці), де домінують синтетична вербалізація (10) та функціональна заміна (6) при мінімальному буквализмі, свідчить про здатність ВММ сприймати ці конструкції як цілісні семантичні блоки. Для дієслів *have* (18 одиниць) та *give* (14 одиниць) характерною рисою стала відносно висока частка перифразу, зумовлена необхідністю описового відтворення станів або процесів передачі інформації, притаманних цим лексемам у медійному дискурсі. Найменш вживане дієслово *get* (7 одиниць) підтверджує загальну тенденцію до синтетизації (3 випадки), хоча його низька частотність, ймовірно, пояс-

нюється стилістичними обмеженнями жанру Lifestyle. Узагальнюючи, можна стверджувати, що рівень успішності подолання структурного розриву між мовами корелює з частотністю вживання десемантизованого дієслова: чим стабільнішою є конструкція в англійській мові, тим ефективніше алгоритми ШІ інтегрують її в синтетичну систему української мови.

Висновки. Проведений корпусний аналіз перекладів медійних текстів рубрики Lifestyle видання «The Guardian», здійснених за допомогою генеративного машинного перекладу (чат GPT), дає підстави встановити наступні закономірності. Домінантною стратегією відтворення англійських аналітичних конструкцій в українському медіа-дискурсі є синтетична вербалізація (45%). Це підтверджує здатність ВММ ідентифікувати аналітичні предикати як семантично цілісні одиниці та успішно трансформувати їх у синтетичні форми, що відповідає типологічним нормам української мови та вимогам динамічності медійного викладу. Другою за частотністю стратегією визначено функціональну заміну (28%), яка забезпечує контекстуально вмотивований добір дієслів, притаманних українській медійній комунікації, тоді як перифраз (17%) слугує інструментом жанрової адаптації для розгортання семантики відповідно до очікувань аудиторії. Обмеження у використанні ВММ виявляються у буквальному покомпонентному перекладі (10%), що є результатом нерозпізнання аналітичного статусу конструкції.

Результати дослідження засвідчують, що успішність нівелювання міжмовної асиметрії ВММ безпосередньо залежить від ступеня десемантизації англійських дієслів. Конструкції з лексемами *take* та *have* демонструють найвищу схильність до синтетизації. Натомість сполуки з *make* та *do* виявляють значну варіативність перекладацьких рішень, хоча саме для *make* характерний прояв «нездоланої» асиметрії у формі буквализ-

Таблиця 1

Кількісні показники перекладацьких стратегій подолання асиметрії КДД за типами дієслів

Десемантизоване дієслово	Синтетична вербалізація	Функціональна заміна	Перифраз	Буквальний переклад	Всього
do	13	8	6	3	30
make	16	11	5	4	36
take	10	6	4	2	22
give	7	5	1	1	14
have	9	4	3	2	18
get	3	1	2	1	7
Всього	58	35	21	13	127

мів. Використання перифразу та функціональної заміни для дієслів *give* та *get* підтверджує намагання алгоритмів ШІ компенсувати структурну невідповідність задля збереження прагматичного потенціалу медіатексту. Загалом результати підтверджують, що ВММ демонструють позитивну тенденцію до подолання структурно-семантичної асиметрії, проте їхня «перекладацька поведінка»

залишається залежною від структурної прозорості КДД та специфіки вхідного контексту.

Перспективи подальших розвідок вбачаються у поглибленні корпусного виміру через залучення процедур постредагування, що дозволить оцінити ефективність подолання мовної асиметрії в медіатекстах в умовах синергії генеративного ШІ та професійної компетенції редактора»

Список літератури:

1. Бойченко Н. О. Стійкі дієслівні сполуки у публіцистичному тексті: типологічні ознаки та експресивний потенціал : дис. канд. філол. наук. КНУТШ, 2001. 239 с.
2. Вихованець І. Р. Дієслівно-іменниковий граматичний тип української мови. *Українська мова*. 2012. № 2. С. 3–10.
3. Городенська К. Морфологічні категорії дієслова у традиційних і нових Зсемантико-граматичних вимірах. *Мовознавство*. 2024. № 1. С. 39–47. DOI: <https://doi.org/10.33190/0027-2833-334-2024-1-004>
4. Загнітко А. Мовний простір граматики : [монографія]. Вінниця: Нілан-Лтд, 2018. 448 с.
5. Іваницька Н. Англійські аналітичні дієслівні лексеми: перекладознавчий аспект. *Наукові записки. Серія: Філологічні науки*. 2023. Вип. 154. С. 268–272.
6. Іваницька Н. Сучасна кваліфікація аналітичних дієслів у аспекті українсько-англійського зіставлення. *Наукові записки Вінницького державного педагогічного університету імені Михайла Коцюбинського. Серія: Філологія (мовознавство): зб. наук. праць / наук. ред. Н. Л. Іваницька*. Вінниця, 2011. Вип. 16. С. 56–61.
7. Іваницька Н. Б., Терещенко Л. Я. Корпусний підхід до виявлення фокалізаторів у романі Айріс Мердок «Пора Янголів». *Сучасні дослідження з іноземної філології*. 2025. №1 (127). С. 285–295. <https://doi.org/10.32782/2617-3921.2025.27.285-295>
8. Романюха М. В. Аспекти перекладу дієслова у англомовних науково-популярних текстах. *Науковий вісник Міжнародного гуманітарного університету. Сер.: Філологія*. 2023. № 61, Том 1. С. 227–231. DOI: <https://doi.org/10.32782/2409-1154.2023.61.50>
9. Robert-Foley L. *Experimental Translation: The Work of Translation in the Age of Algorithmic Production*. Goldsmith Press, 2024. 254 p.
10. Asscher Omri. The position of machine translation in translation studies. A definitional perspective. *Translation Spaces*. 2023. Volume 12, Issue 1. P. 1-20. DOI: <https://doi.org/10.1075/ts.22035.ass>
11. Bruening B. Light verbs are just regular verbs. *University of Pennsylvania Working Papers in Linguistics*. 2015. Vol. 21, № 1. P. 51–60.
12. Giparaitė J. A corpus-based analysis of light verb constructions with MAKE and DO in British English. *Kalbotyra*. 2024. № 76. P. 18–41. DOI: <https://doi.org/10.15388/Kalbotyra.2023.76.2>
13. Leech G. *A Glossary of English Grammar*. Edinburgh: Edinburgh University Press, 2006. 384 p.
14. Martínez Caro E., Arús-Hita J. Give as a light verb. *Functions of Language*. 2020. Vol. 27, № 3. P. 280–306. DOI: <https://doi.org/10.1075/FOL.16036.MAR>
15. Mehl S. Light verb semantics in the International Corpus of English: Onomasiological variation, identity evidence, and degrees of lightness. *English Language and Linguistics*. 2019. Vol. 23, № 1. P. 55–80. DOI: <https://doi.org/10.1017/S1360674317000302>
16. Musayeva N. The Role of Light Verbs in English Syntax: Crosslinguistic Perspectives and Structural Implications. *EuroGlobal Journal of Linguistics and Language Education*. 2024. 1(2). P. 100–108. DOI:10.69760/3yvcx94
17. Riccio A., Fleischhauer J., eds. *Light verbs: Synchronic and diachronic studies*. Berlin/Boston: Walter de Gruyter GmbH, 2025. 286 p. DOI: <https://doi.org/10.1515/9783111388878>.
18. Ronan P., Schneider G. Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics*. 2015. Vol. 20, № 3. P. 326–354. DOI: 10.1075/ijcl.20.3.03ron
19. Wittenberg E. *With Light Verb Constructions from Syntax to Concepts*. Potsdam: Universitätsverlag Potsdam, 2016.

Ivanytska N. B. CORPUS-BASED ANALYSIS OF LINGUISTIC ASYMMETRY IN LIGHT VERB CONSTRUCTIONS IN GPT-GENERATED MEDIA TRANSLATION

The article presents the results of a comprehensive corpus-based study of linguistic asymmetry in the translation of delexicalized verbs in media texts by large language models. The primary focus is on the specific functioning and rendering of English light verb constructions (LVCs), particularly productive units such as do, make, take, give, and have, within the framework of English-Ukrainian machine translation. The research material consists of a specially developed parallel corpus including original lifestyle publications from the British media outlet The Guardian and their direct translations generated by the GPT large language model. The relevance of the scientific inquiry is driven by the pronounced structural and semantic asymmetry in English LVCs, in which lexical meaning is concentrated in the nominal component. At the same time, the verb performs a predominantly grammatical function, and the synthetic structure of the Ukrainian language tends toward the use of full-valued verbal predicates. This interlinguistic discrepancy poses specific cognitive and algorithmic challenges for automated translation systems, requiring a detailed study of the LLM's ability to adapt these constructions to the target-language norms. The methodological framework of the study is based on the application of Sketch Engine professional tools, which enabled precise alignment of parallel corpus segments, automated extraction of relevant constructions, and subsequent classification by translation strategies. The scientific sample comprises 127 verified contexts selected according to the presence of an abstract nominal component, the possibility of paraphrasing with a simple verb, and high collocational stability in media discourse. The results of the corpus analysis indicate that synthesis serves as the dominant strategy for overcoming asymmetry (≈46% of cases). This demonstrates a high level of LLM capability in integrating a delexicalized verb and a noun into a single synthetic predicate in Ukrainian, ensuring pragmatic accuracy and stylistic consistency in the media text. Additionally, the paper provides a detailed description of the application of functional substitution and paraphrase (explication) strategies, while documenting cases of literal, component-by-component translation that represent a risk to machine translation quality. The obtained data confirm the hypothesis of the evolving adaptability of large language models to complex cases of linguistic asymmetry in media discourse.

Keywords: *light (delexicalized) verbs, English–Ukrainian generative machine translation, large language model, parallel corpus, media discourse, translation strategies, corpus analysis, analytic constructions, interlingual asymmetry.*

Дата першого надходження статті до видання: 24.01.2026

Дата прийняття статті до друку після рецензування: 23.02.2026

Дата публікації (оприлюднення) статті: 24.04.2026