



УДК УДК 81'33

[https://doi.org/10.52058/2786-6165-2025-2\(32\)-494-507](https://doi.org/10.52058/2786-6165-2025-2(32)-494-507)

Саліонович Людмила Миколаївна кандидат філологічних наук, доцент кафедри міжкультурної комунікації та іноземної мови, Національний технічний університет «Харківський політехнічний інститут», м. Харків, вул. Кирпичова 2, тел.: (057) 707-69-70, <https://orcid.org/0000-0001-9005-9273>

Рубцова Вікторія Вікторівна доцент кафедри міжкультурної комунікації та іноземної мови, Національний технічний університет «Харківський політехнічний інститут», м. Харків, тел.: (057) 707-69-70, <https://orcid.org/0000-0003-1120-5514>

Нетецька Тетяна Миколаївна доцент кафедри міжкультурної комунікації та іноземної мови, Національний технічний університет «Харківський політехнічний інститут», м. Харків, тел.: (057) 707-69-70, <https://orcid.org/0000-0001-5298-0563>

ЕВОЛЮЦІЯ ЕКОНОМІЧНОЇ ТЕРМІНОЛОГІЇ: «ТРИЛОГІЯ БАЖАННЯ» ТЕОДОРА ДРАЙЗЕРА VS СУЧАСНИЙ ДИСКУРС

Анотація. У статті досліджено та проаналізовано еволюцію економічних термінів, що відбулася за понад сто років, на прикладі трилогії Теодора Драйзера та сучасного дискурсу. За основу методики дослідження були взяті корпусні технології, за допомогою яких з текстів трилогії були виокремлені іменники, що являють собою центральні економічні концепти і мають найбільшу контекстуальну вагу, що визначається на основі методу TF-IDF. Дана технологія враховує не лише частотність вживання слова, а й високу специфічність у контексті корпусу, на основі чого проводиться порівняльний аналіз термінів на предмет їх функціонування у сучасній англійській мові. На першому етапі дослідження було виокремлено корпус з 200 найбільш важливих слів, тобто слів з високими вагами. З отриманого переліку за експертною оцінкою було виділено 50 слів, що відносяться до економічної лексики. Другий етап дослідження полягав у порівнянні частотності використання економічних термінів у корпусах трилогії Теодора Драйзера та сучасної англійської мови. Враховувалися три показники: чисто входжень, частота на мільйон токенів та відсоток від корпусу. Отримані показники дозволили виявити відмінності вживання термінів у художньому тексті та спеціальному економічному підкорпусі,



створеному на базі програми English Trends, особливості їх функціонування тощо. Аналіз виявив зміщення акцентів у деяких термінах, що є неминучим наслідком розвитку цифрових технологій, колективного інвестування та глобалізації фінансових ринків. Діахронічний аналіз виявив також стійкість деяких економічних термінів. Було виявлено особливості функціонування економічної термінології у різних типах дискурсу шляхом зіставлення художнього та спеціалізованого корпусів. Перспектива подальших досліджень вбачається у аналізі словосполучень, що включають економічні терміни.

Ключові слова: корпус, економічна термінологія, лема, токен, частотність вживання, вилучення термінів.

Salionovych Liudmyla Mykolayivna PhD in Philology, Associate Professor of the Department of Cross-cultural Communication and Foreign Languages, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, tel.: (057) 707-69-70, <https://orcid.org/0000-0001-9005-9273>

Rubtsova Victoria Victorivna Associate Professor of the Department of Cross-cultural Communication and Foreign Languages, National Technical University "Kharkiv Polytechnic Institute", tel.: (057) 707-69-70, <https://orcid.org/0000-0003-1120-5514>

Netetska Tetiana Mykolaiivna Associate Professor of the Department of Cross-cultural Communication and Foreign Languages, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, tel.: (057) 707-69-70, <https://orcid.org/0000-0001-5298-0563>

EVOLUTION OF ECONOMIC TERMINOLOGY: THEODORE DREISER'S "TRILOGY OF DESIRE" VS MODERN DISCOURSE

Abstract. The paper represents and analyzes the evolution of economic terms that has occurred over a hundred years, using the texts of Theodore Dreiser's trilogy and modern discourse. The research methodology was based on corpus technologies, with the help of which nouns that represent central economic concepts and have the greatest contextual weight, determined based on the TF-IDF method, were singled out from the texts of the trilogy. This technology takes into account not only the frequency of use of a word, but also high specificity in the context of the corpus, on the basis of which a comparative analysis of terms is carried out for their functioning in modern English. At the first stage of the study, a corpus of 200 of the most important words was singled out, i.e. words



with high weights. From the resulting list, 50 words related to economic vocabulary were selected by expert assessment. The second stage of the study consisted of comparing the frequency of use of economic terms in the corpora of Theodore Dreiser's trilogy and modern English. Three indicators were taken into account: pure occurrences, frequency per million tokens and percentage of the corpus. The obtained indicators made it possible to identify differences in the use of terms in the literary text and the special economic subcorpus created on the basis of the English Trends program, as well as the features of their functioning. The analysis revealed a shift in emphasis in some terms, which is an inevitable consequence of the development of digital technologies, collective investment and the globalization of financial markets. Diachronic analysis also revealed the stability of some economic terms. The features of the functioning of economic terminology in different types of discourse were revealed by comparing the literary and specialized corpora. The prospect of further research is seen in the analysis of word combinations that include economic terms.

Keywords: corpus, economic terminology, lemma, token, frequency of use, terminology extraction.

Постановка проблеми. Корпусні технології відіграють ключову роль у сучасних лінгвістичних дослідженнях, надаючи системні підходи до аналізу, вилучення та управління термінами на основі великих текстових колекцій, відомих як корпуси. Корпус – універсальний інструмент для вивчення мови, оскільки являє собою велику машинно-читану і часто анотовану для зручності аналізу, структуровану колекцію текстів або мовлення, призначену для представлення певної мови або її різновиду, що використовується для лінгвістичного аналізу та досліджень. Цей ресурс став невід'ємною частиною сучасних лінгвістичних досліджень, особливо в таких галузях, як машинний переклад, управління термінологією, лексикографія та термінографія. У контексті лінгвістичних досліджень корпус може використовуватися для статистичного тестування гіпотез та перевірки лінгвістичних правил. Це особливо важливо для таких областей, як машинний переклад, де потрібен аналіз великих обсягів даних та термінологія, де необхідне вилучення спеціалізованих термінів.

Корпус має низку переваг, які роблять його цінним ресурсом для різних лінгвістичних завдань. Однією з ключових переваг корпусу є його здатність надавати автентичні дані, що відображають реальне використання мови, що дозволяє аналізувати мову в її природному контексті. Це особливо важливо для машинного перекладу, де паралельні корпуси, які включають тексти та його переклади, використовуються для навчання моделей. У термінознавстві такі дані допомагають витягувати терміни з реальних



текстів, що полегшує управління термінологією. Корпус дозволяє проводити як кількісний, так і якісний аналіз, що допомагає виявляти мовні патерни та тенденції. Великі колекції текстів дозволяють проводити статистичний аналіз, який важко реалізувати якісними методами. Наприклад, у лексикографії аналіз частоти та колокацій допомагає створювати словники, що відображають реальне використання слів. У термінографії це допомагає визначати значимість термінів у певних областях.

Аналіз останніх досліджень і публікацій. Корпусна лінгвістика є одним із ключових напрямів сучасних мовознавчих досліджень, що поєднує методи автоматизованої обробки текстів, статистичного аналізу та цифрового моделювання мовних явищ. Огляд наукових статей, присвячених цій тематиці, демонструє широке коло застосувань корпусних методів, від лексикографії та термінографії до когнітивного аналізу й діахронічного дослідження спеціалізованої лексики. Представлені дослідження охоплюють різні аспекти корпусної лінгвістики: створення та використання корпусів для аналізу термінології в економічній, ветеринарній, морській та офіційно-діловій сферах, а також вивчення мовних особливостей художніх і фольклорних текстів. Особливий інтерес становлять роботи, що демонструють застосування спеціалізованого програмного забезпечення, такого як CLAWS, AntConc, Sketch Engine та WordSmith Tools, для морфологічної, семантичної й стилістичної розмітки корпусів. Огляд досліджень засвідчує зростаюче значення корпусного аналізу для розвитку лінгвістичних і прикладних дисциплін, підтверджуючи ефективність цифрових технологій у вивченні та систематизації мовних даних.

Стаття Н. В. Нікуліної [1] аналізує сучасні наукові проекти в галузі комп'ютерної лінгвістики в Україні, зосереджені на роботі з мовними корпусами. У ній розглядається використання цифрових технологій, таких як автоматизована обробка текстів, статистичний аналіз і створення машиночитабельних баз даних, для дослідження мови. Основна ідея статті полягає в тому, щоб показати, як ці технології, включно з вебінтерфейсами, алгоритмами класифікації та інструментами пошуку, сприяють науковим дослідженням і покращують методи викладання. Проекти демонструють важливість цифрових підходів для аналізу текстів і їхній внесок у глобальну співпрацю дослідників.

Дослідження Д. Ю. Бабурніча та Є. В. Купріянова [2] зосереджене на аналізі спеціалізованих корпусів текстів і їхній ролі в науково-технічній лексикографії. Авторами розглядається процес створення корпусу текстів з екологічної економіки для потреб лексикографів у складанні термінологічного словника. Визначається структура корпусу, його наповнення та методи обробки, такі як морфологічна розмітка за допомогою системи

CLAWS C7. Запропоноване дослідження підкреслює важливість структурованих текстових баз даних для автоматичної обробки текстів, а також демонструє можливості цифрових технологій у створенні лексикографічних ресурсів.

Ю. Г. Рожков [3] у своїй статті застосовує корпусні методи для аналізу термінології ветеринарної медицини. Дослідження поєднує когнітивний підхід та принципи корпусної лінгвістики для вивчення термінів, що використовуються у ветеринарній сфері. Використовуючи корпус текстів із Black's Veterinary Dictionary, автор застосовує методи автоматичного виділення термінів, включно з кластерним аналізом і статистичними інструментами. Це дозволяє сформулювати список термінів для онтології ветеринарної галузі та забезпечує ефективну систематизацію спеціалізованих знань.

Стаття Т. П. Дяк і Ю. І. Грицюк [4] розглядає використання корпусних інструментів AntConc та Sketch Engine для аналізу текстів фольклорного дискурсу, зокрема стрілецьких і повстанських пісень. Дослідники демонструють ефективність цих інструментів для частотного аналізу, пошуку колокацій та термінологічних сполучень. Використання таких функцій, як Word List, Collocates, N-Grams і Keywords, дозволяє ідентифікувати ключові слова й лінгвістичні особливості текстів. Це дослідження підтверджує потенціал корпусних технологій у вивченні й систематизації усної народної творчості.

О. В. Цобенко [5] у своєму дослідженні вивчає діахронічний аспект англійської морської термінології. Використовуючи спеціалізовані словники та етимологічний аналіз, автор простежує зміни в терміносистемі через зіставлення історичних мовних контактів і запозичень. Дослідження спирається на текстові масиви, які відображають технологічні й соціокультурні зміни в морській справі. Аналізуючи лексичні зміни у різні історичні періоди, автор демонструє вплив міжнародних стандартів на сучасну морську термінологію.

О. О. Жулавська та А. О. Дорогокупля [6] аналізують використання корпусних менеджерів WordSmith Tools і Sketch Engine для дослідження безособових форм дієслів. У статті розглядаються методи лексичного та граматичного аналізу текстів офіційно-ділового стилю, зокрема з перекладацької перспективи. Дослідження демонструє, що перекладацькі трансформації є необхідними для точності передачі значення між англійською та українською мовами. Корпусні технології дозволяють перекладачам створювати власні текстові корпуси, що допомагає уточнювати частотність вживання лексем та їхню відповідність у конкретному контексті.

За кордоном проводяться масштабні дослідження щодо технологій вилучення термінів із текстів. Статті, присвячені автоматичному вилученню термінології, охоплюють широкий спектр підходів і технологій у



галузі обробки природної мови (NLP), демонструючи еволюцію методів від простих статистичних до складних семантичних рішень. Досліджується, як із зростанням обсягу текстів у цифровому просторі, зокрема через соціальні медіа та великі дані, аналіз термінології стає ключовим для таких завдань, як машинний переклад, побудова онтологій і організація знань, простежуючи розвиток від традиційних лінгвістичних і статистичних методів до гібридних, графових, машинного та глибокого навчання, а також сучасних мовних моделей, описується типова архітектура систем вилучення термінів (попередня обробка, відбір кандидатів, ранжування), аналізуються набори даних, інструменти й метрики оцінки [7]. Увага дослідників зосереджена на вилученні термінів із документів, зазначаючи обмеження частотних методів через специфіку й малий обсяг текстів, розроблюються методи, які інтегрують частотний аналіз із вкладеннями слів (word embeddings), аналізуючи контекст для ідентифікації термінів [8]. Пропонується метод C-value/NC-value для розпізнавання багатослівних термінів, комбінуючи лінгвістичні й статистичні принципи: C-value оцінює частоту й довжину словосполучень, а NC-value додає контекстний аналіз для точності [9].

Таким чином, аналіз представлених досліджень свідчить про значну роль корпусної лінгвістики у вивченні спеціалізованих терміносистем, аналізі текстових структур та розвитку методів автоматизованої обробки мовних даних. Запропоновані в дослідженнях підходи демонструють ефективність цифрових технологій у розв'язанні лінгвістичних задач, що підтверджує важливість подальших досліджень у цьому напрямі.

Мета статті. Метою нашого дослідження, реалізованого за допомогою коду мовою Python, є ідентифікація, а також кількісний і якісний аналіз термінологічного складу текстового корпусу, створеного нами на матеріалі трьох романів «Трилогії бажання» Теодора Драйзера (The Trilogy of Desire) з метою виявлення іменників, які імовірно являють собою центральні економічні концепти в тексті романів Драйзера і мають найбільшу контекстуальну значущість, або інакше кажучи вагу, що визначається на основі методу TF-IDF (Term Frequency-Inverse Document Frequency) із застосуванням логарифмічного згладжування частоти для виділення слів-термінів, які не просто часто зустрічаються в тексті, а мають високу специфічність у контексті корпусу, що відображає їх роль у формуванні економічного наративу автора, на основі чого проводиться порівняння отриманих однослівних термінів із їх функціонуванням у сучасній мові на матеріалі великого репрезентативного корпусу англійської мови English Trends (економічний підкорпус).

Виклад основного матеріалу. Розглянемо коротко функції основних інструментів, які традиційно використовуються для обробки тексту

природною мовою. Спеціалізовані бібліотеки spaCy застосовуються для обробки природної мови, включаючи розбір речень, визначення частин мови та виділення іменованих сутностей; NLTK надає інструменти для роботи з текстами, включаючи зумовлені списки стоп-слів, токенізацію та лематизацію; string використовується для роботи з рядками та наборами символів, включаючи видалення пунктуації; math необхідний для виконання математичних обчислень, таких як логарифмічне згладжування. Counter з модуля collections використовується для роботи з частотними розподілами слів, що важливо для обчислення термінологічної значущості. Tf/idfVectorizer із sklearn.feature_extraction.text дозволяє застосовувати метод TF-IDF, який оцінює значущість слів у тексті. Модуль nltk.download("stopwords") завантажує набір стоп-слів, включених до бібліотеки NLTK. Ці слова являють собою елементи, що найчастіше зустрічаються в мові (наприклад, "the", "is", "and"), які не несуть значного смислового навантаження і тому підлягають видаленню в процесі аналізу тексту. Таким чином, виключення нерелевантних слів здійснюється для покращення якості подальшої обробки тексту та виділення ключових термінів. Цей етап критично важливий для фільтрації шуму у тексті, оскільки стоп-слова можуть спотворювати результати аналізу. Відбувається аналіз іменованих сутностей (Named Entity Recognition, NER), і з тексту витягуються лише ті, що мають мітку "PERSON" (імена власних людей) або "GPE" (геополітичні об'єкти, такі як міста, країни, регіони). Створюється безліч іменованих сутностей, в якій всі слова приводяться до нижнього регістру, щоб уникнути дублювання (наприклад, "New York" і "new york" вважатимуться одним і тим самим об'єктом). Цей етап допомагає видалити зайві елементи тексту, які можуть впливати на аналіз, але не несуть смислового навантаження у контексті виділення ключових термінів. Лематизація та токенізація є фундаментальними етапами обробки тексту у завданнях аналізу природної мови. У процесі токенізації текст розбивається на окремі одиниці (токени), такі як слова чи фрази. Цей процес є важливим для подальшої обробки, оскільки дозволяє проводити аналіз на рівні окремих слів. Без токенізації неможливо ефективно застосовувати методи машинної обробки, такі як TF-IDF, оскільки вони потребують роботи з дискретними елементами тексту. У ході лематизації слів вони приводяться до їхньої базової форми, званої лемою. На відміну від стемінгу, який просто відсікає закінчення, лематизація враховує граматичні та морфологічні характеристики слова, що робить її точнішою. Наприклад, слова "running", "ran" та "runs" будуть приведені до однієї лемі "run". Це знижує надмірність даних, покращує якість пошуку збільшує точність аналізу тексту.

Перший етап дослідження було спрямовано на вилучення ключових іменників з корпусу трилогії, що відбивають економічну семантику тексту,



з наступним ранжуванням їх за значимістю. Такий підхід дозволяє встановити найбільш релевантні лексичні одиниці, які формують економічний дискурс у художній розповіді, виключаючи вплив власних і загальноновживаних стоп-слів. Слід зазначити, що в області корпусної лінгвістики та комп'ютерної обробки текстів аналіз ваги слів із застосуванням TF-IDF є визнаним методом для визначення лексичної значущості у великих текстових масивах. В даному випадку дослідження зосереджено на іменниках як базових носіях концептуальної інформації в економічному контексті, що обумовлено їхньою ключовою роллю в термінології фінансової сфери. Трилогія Драйзера, що відображає реалії американського капіталізму початку ХХ століття, є багатим матеріалом для вивчення економічної лексики в художній літературі. Текст трилогії насичений економічною термінологією, що відбиває прагнення автора до граничного реалізму зображення світу фінансового капіталу кінця ХІХ століття. Його романи відтворюють не лише події, а й мову, якою оперують біржові спекулянти, банкіри та підприємці. Економічна лексика стає невід'ємною частиною художнього методу письменника, дозволяючи створити достовірну картину фінансових механізмів, які управляють долями людей. Однак використання специфічної термінології у трилогії не зводиться лише до відтворення професійного середовища головного героя. Воно виконує важливу смислову функцію, розкриваючи глибинні закономірності капіталістичної системи, у якій людські амбіції, моральні принципи та соціальні відносини підпорядковуються логіці грошового обігу. Драйзер демонструє, що фінансовий ринок – це не лише простір економічних операцій, а й арена соціальних взаємодій, де виграє той, хто краще розуміє його правила. Економічні терміни в цьому контексті перестають бути суто професійними маркерами і перетворюються на символи влади, успіху, маніпуляції, а часом неминучого краху. Мислення головного героя організовано навколо цифр, котирувань, угод та балансів, на чому наголошується завдяки відповідній термінології. Через мову економічної сфери автор розкриває раціональність, прагматизм персонажа та здатність вбачати у грошах інструмент влади. Тим самим Драйзер не просто визначає становлення фінансового генія, а й розкриває психологічні механізми підприємницького успіху. Таким чином, мова економіки в романі виконує подвійну функцію: з одного боку, вона занурює читача у світ фондових бірж і банків, з іншого – стає інструментом критичного аналізу соціально-економічних відносин. Враховуючи описану вище значимість термінів і виконуване ними смислове навантаження в романах, метою нашого дослідження є не просто виділення частотних слів, а ідентифікація тих іменників, які мають високу специфічність і

контекстуальну значущість, що досягається за рахунок виключення нетермінологічних елементів (стоп-слів та власних назв) і застосування статистичного методу оцінки їх ваги.

Створений нами код на мові програмування Python реалізує багатоетапний підхід до аналізу текстового корпусу, поєднуючи інструменти обробки природної мови (NLP) та статистичні методи. Основні етапи включають попередню обробку тексту з використанням SpaCy для морфологічного аналізу, що дозволяє лематизувати текст і фільтрувати тільки іменники (`token.pos_ == "NOUN"`), виключаючи інші частини мови, що в свою чергу забезпечує фокус на термінологічній лексиці, характерній для економічного дискурсу, а також до цього етапу віднесемо видалення стоп-слів з NLTK та вилучених власних назв (PERSON GPE) за допомогою функції `extra_ct_named_entities`, що мінімізує шум від загальноживаних слів та контекстно-специфічних імен, таких як назви місць чи персонажів; сегментацію корпусу, що передбачає розбиття тексту на абзаци (`corpus = text.split("\n\n")`), які розглядаються як окремі документи, це дозволяє враховувати розподіл термінів за різними контекстами всередині корпусу, посилюючи точність IDF-компоненти; обчислення TF-IDF з логарифмічним згладжуванням з функцією `compute_tf`, яка застосовує логарифмічне згладжування частоти термінів (`1 + math.log10(count)`), що зменшує вплив екстремально частих слів і підкреслює значущість помірно-частотних термінів, при цьому `TfidfVectorizer` з бібліотеки `scikit-learn` обчислює підсумкові ваги, комбінуючи TF та IDF, де IDF відображає рідкість слова в корпусі, посилюючи вагу специфічних термінів; і, нарешті етап ранжування та виведення результатів, на якому підсумовуються TF-IDF ваги іменників по всіх документах (`tfidf_matrix.sum(axis=0).A1`), що дозволяє оцінити загальну значущість кожного іменника в корпусі, а сортування за зменшенням ваги виділяє топ-200 найбільш важливих важливих слів. Таким чином була створена основа для подальшого аналізу, оскільки отриманий список топ-200 іменників з високими вагами є емпіричною базою для подальших досліджень, таких як семантичний аналіз, порівняння з іншими текстами або вивчення еволюції термінології. З отриманого списку за допомогою експертної оцінки було виділено 50 слів, які відносяться до економічної лексики, що становить 25 відсотків від усіх слів з високими вагами, що ще раз говорить про високу щільність економічної термінології у тексті, що досліджується. З цих 50 слів було обрано 20 слів із найбільшою кількістю входів.

На другому етапі дослідження ми порівняли частотність використання економічних термінів в обох корпусах. Для аналізу будемо враховувати три ключові показники: число входжень (Number of hits), тобто



абсолютну кількість згадок кожного терміна в корпусі; частота на мільйон токенів (Hits per million tokens), тобто нормалізований показник, що дозволяє порівнювати частотність термінів між корпусами різного розміру; відсоток від корпусу (Percent of whole corpus), або відносна частка терміна в загальному обсязі токенів, виражена у відсотках. Результати порівняння показано в Таблиці 1, де у колонці Lemma (словникові форми) представлені однослівні економічні терміни. Ці показники допомогли виявити відмінності у вживанні термінів у художньому тексті («Трилогія бажання» Драйзера) та спеціалізованому економічному підкорпусі, створеному на базі English Trends, а також визначити закономірності їхнього функціонування.

Таблиця 1

Порівняльний аналіз частотності однослівних економічних термінів в корпусах Trilogy та English Trends

Lemma	Trilogy Number of hits	Trilogy Number of hits per million tokens	Trilogy Percent of whole corpus	English Trends Number of hits	English Trends Number of hits per million tokens	English Trends Percent of whole corpus
interest	346	569.91	0.05699	4,071,471	483.83	0.004151
stock	282	464.49	0.04645	8,801,095	1,045.87	0.008972
loan	239	393.67	0.03937	2,864,011	340.34	0.002920
value	117	192.72	0.01927	4,212,004	500.53	0.004294
check	109	179.54	0.01795	493,959	58.7	0.0005035
deal	109	179.54	0.01795	2,800,260	332.77	0.002855
property	99	163.07	0.01631	2,493,967	296.37	0.002542
franchise	140	230.6	0.02306	379,653	45.12	0.0003870
paper	173	284.96	0.02850	685,944	81.51	0.0006993
treasury	80	131.77	0.01318	666,47	79.2	0.0006794
exchange	66	108.71	0.01087	1,998,626	237.51	0.002037
note	71	116.95	0.01169	1,424,724	169.31	0.001452
broker	101	166.36	0.01664	380,022	45.16	0.0003874
fund	30	49.41	0.004941	4,055,806	481.97	0.004135
account	58	95.53	0.009553	2,445,818	290.65	0.002493
bill	57	93.89	0.009389	1,345,649	159.91	0.001372
share	139	228.95	0.02290	9,859,287	1,171.62	0.01005
security	62	102.12	0.01021	2,979,175	354.03	0.003037
bond	59	97.18	0.009718	1,654,795	196.65	0.001687
certificate	101	166.36	0.01664	199,91	23.76	0.0002038

Розмір корпусу "Драйзер" значно менше, ніж сучасний корпус English Trends (607113 проти 98095378708 слів), що необхідно враховувати при порівнянні абсолютних частот. Для коректнішого порівняння слід орієнтуватися на відносну частотність (на мільйон слів). Як видно з Таблиці 1, найбільш частотними термінами у «Трилогії бажання» є:

"interest" (відсотки) – 569.91 на мільйон слів,
"stock" (акції) – 464.49 на мільйон слів,
"loan" (позика) - 393.67 на мільйон слів,
"paper" (цінні папери) – 284.96 на мільйон слів,
"franchise" (франшиза) - 230.6 на мільйон слів,
"share" (частка, акція) – 228.95 на мільйон слів, "value" (вартість) – 192.72 на мільйон слів.

А для корпусу English Trends найбільш частотні терміни такі:
"share" (частка, акція) – 1,171.62 на мільйон слів,
"stock" (акції) – 1,045.87 на мільйон слів,
"value" (вартість) – 500.53 на мільйон слів,
"interest" (відсотки) – 483.83 на мільйон слів,
"fund" (фонд) – 481.97 на мільйон слів,
"security" (цінний папір) – 354.03 на мільйон слів, "loan" (позика) - 340.34 на мільйон слів.

Зі статистичних даних видно, що терміни "interest", "stock", "loan", "value", "share" зберігають високу частотність в обох корпусах, що свідчить про їх фундаментальну роль в економічній лексиці.

В English Trends спостерігається значне зростання частотності термінів "share", "stock", "fund", що може бути пов'язане з розвитком фондового ринку та інвестиційних інститутів.

Терміни "check", "franchise", "paper", "treasury", "broker", "certificate" мають відносно високу частотність у романах, але значно меншу в English Trends, що може відображати зміни від конкретних економічних реалій епохи Драйзера до сьогодення.

Аналіз частотності та вживання економічних термінів у «Трилогії бажання» Теодора Драйзера порівняно з їхньою присутністю у сучасному економічному підкорпусі English Trends виявляє значні діахронічні зміни у фінансовій лексиці. Цей аналіз не лише підкреслює еволюцію економічної мови, а й відбиває ширші соціально-економічні трансформації, які впливають на мову та її використання з часом.

У романах Драйзера деякі терміни зустрічаються набагато частіше, ніж у сучасних текстах, що свідчить про характерні риси фінансового дискурсу кінця XIX – початку XX століття.

Наприклад, слово *franchise* у творах Драйзера означає права чи привілеї, що надаються державою чи корпорацією, особливо у таких



сферах, як комунальні послуги та транспорт. Сьогодні цей термін найчастіше використовується у значенні бізнес-моделі, за якої незалежні підприємці отримують ліцензію на використання бренду та бізнес-процесів великої компанії.

Термін *broker* у творах Драйзера означає посередника на фінансовому ринку, який купує та продає цінні папери від імені клієнтів. У сучасному контексті роль брокерів значно змінилася через поширення онлайн-трейдингу, який знизив потребу у традиційних посередниках.

Аналогічно, слово *certificate* у текстах Драйзера відноситься до фізичного документа, що підтверджує володіння акціями або облігаціями, тоді як сьогодні більшість таких сертифікатів існують у цифровій формі, а сам термін може також означати електронні сертифікати безпеки веб-сайтів.

Інший примітний термін – *paper*. У романах Драйзера він часто використовується у значенні цінних паперів чи фінансових документів, тоді як у сучасній мові слово *paper* найчастіше вживається відносно наукових публікацій чи газетних статей.

Термін *treasury* в ХІХ столітті в основному означав державні фінанси або відомство казначейства, але в сучасних текстах його вживання змістилося у бік позначення казначейських облігацій та державних цінних паперів.

Деякі терміни, навпаки, стали більш поширеними у сучасних текстах, ніж у романах Драйзера.

Наприклад, слово *stock*, що означає акції компаній, сьогодні зустрічається набагато частіше, що пов'язано зі зростанням популярності інвестицій та активної участі населення у фондовому ринку.

Аналогічно, термін *fund* набув широкого розповсюдження завдяки появі та розвитку взаємних фондів, хедж-фондів та інших колективних інвестиційних інструментів.

Термін *share*, що означає частку володіння у компанії, також використовується частіше, що відображає демократизацію інвестування та зростаюче значення акціонерної вартості у корпоративному управлінні.

Слово *security* традиційно використовувалося у значенні фінансових інструментів, таких як акції та облігації, але в сучасній мові його значення розширилося, включивши також заходи кібербезпеки, що підкреслює технологічну трансформацію фінансової сфери.

Термін *bond* зберіг своє значення і продовжує широко використовуватися, оскільки боргові зобов'язання залишаються ключовим елементом фінансування як корпорацій, так і державних структур.

Висновки. Таким чином, аналіз показує, що економічна лексика, що використовується Драйзером, відображає особливості фінансової системи

його епохи, тоді як сучасний економічний дискурс демонструє зміщення акценту на цифрові технології, колективне інвестування та глобалізацію фінансових ринків. Діахронічний аналіз виявив як стійкість, так і мінливість у частотності економічних термінів. Стійкість свідчить про збереження фундаментальних економічних понять. Мінливість відображає еволюцію економічних систем, появу нових фінансових інструментів та інституцій.

Зіставлення художнього та спеціалізованого корпусів дозволяє виявити особливості функціонування економічної термінології у різних типах дискурсу. Дані вказують на те, що економічна термінологія в трилогії Драйзера відображає конкретні економічні реалії своєї епохи, але водночас демонструє наступність сучасної економічної лексики.

Перспективу подальших досліджень вбачаємо в аналізі словосполучень, що включають економічні терміни, а також у вивченні їхньої семантичної еволюції та контекстуальних змін у ширшому діахронічному контексті з використанням сучасних корпусних технологій.

Література:

1. Нікуліна Н. В. Загальний огляд комп'ютерно-мовознавчих наукових проєктів України. Лінгвістичні дослідження: зб. наук. пр. Харк. нац. пед. унту імені Г. С. Сковороди / гол. ред. Н. В. Піддубна. Харків, 2024. Вип. 61. С. 314–322. DOI: <https://doi.org/10.34142/23127546.2024.61.25>

2. Бабурніч Д. Ю. Використання корпусу економічних текстів для розробки тематичного словника / Д. Ю. Бабурніч, Є. В. Купріянов // Вісник Національного технічного університету "ХПІ". Сер. : Актуальні проблеми розвитку українського суспільства = Bulletin of the National Technical University "KhPI". Ser. : Actual problems of Ukrainian society development : зб. наук. пр. – Харків : НТУ "ХПІ", 2020. – № 1. – С. 77-81. DOI: <https://doi.org/10.20998/2227-6890.2020.1.14>

3. Рожков Ю. Г. Методика виокремлення термінів ветеринарної медицини з корпусу ветеринарних текстів / Ю. Г. Рожков // Закарпатські філологічні студії. - 2019. - Вип. 9(2). - С. 33-36. - URL: [http://nbuv.gov.ua/UJRN/trphst_2019_9\(2\)_8](http://nbuv.gov.ua/UJRN/trphst_2019_9(2)_8).

4. Дяк, Т. П., & Грицюк, Ю. І. (2024). Використання корпусних інструментів для виявлення ключових слів у стрілецьких і повстанських піснях як жанру фольклорного дискурсу. Scientific Bulletin of UNFU, 34(7), 60-71. <https://doi.org/10.36930/40340708>

5. Цобенко, О. В. Діахронічний аспект дослідження англословної морської термінології. Закарпатські філологічні студії. 2024. Вип. 36. С. 136-139. URL: <http://zfs-journal.uzhnu.uz.ua/archive/36/25.pdf>

6. Жулавська О.О., Дорогокупля А.О. Дослідження безособових форм дієслів програмними засобами корпусної лінгвістики. Актуальні питання гуманітарних наук: міжвузівський збірник наукових праць молодих вчених Дрогобицького державного педагогічного університету імені Івана Франка / [редактори-упорядники М. Пантюк, А. Душний, І. Зимомря]. Дрогобич: Видавничий дім «Гельветика», Вип. 34. Том 2. 2020. С. 131-139

7. Xu, K., Feng, Y., Li, Q. et al. Survey on terminology extraction from texts. J Big Data 12, 29 (2025). DOI: <https://doi.org/10.1186/s40537-025-01077-x>



8. M. Bay, D. Bruneß, M. Herold, C. Schulze, M. Guckert and M. Minor, "Term Extraction from Medical Documents Using Word Embeddings," *2020 6th IEEE Congress on Information Science and Technology (CiSt)*, Agadir - Essaouira, Morocco, 2020, pp. 328-333, DOI: <https://doi.org/10.1109/CiSt49399.2021.9357263>.

9. Frantzi, K., Ananiadou, S. & Mima, H. Automatic recognition of multi-word terms: the C-value/NC-value method. *Int J Digit Libr* 3, 115–130 (2000). DOI: <https://doi.org/10.1007/s007999900023>

References:

1. Nikulina, N. V. (2024). Zahalnyi ohliad komp'uterno-movoznavchykh naukovykh proektiv Ukrainy [General overview of computer-linguistic scientific projects in Ukraine]. *Linhvistychni doslidzhennia: Zbirnyk naukovykh prats Kharkivskoho natsionalnoho pedahohichnoho universytetu imeni H. S. Skovorody*, (61), 314–322. <https://doi.org/10.34142/23127546.2024.61.25>

2. Baburnich, D. Yu., & Kupriianov, Ye. V. (2020). Vykorystannia korpusu ekonomichnykh tekstiv dlia rozrobky tematicnoho slovnyka [Using a corpus of economic texts for developing a thematic dictionary]. *Visnyk Natsionalnoho tekhnichnoho universytetu "KhPI". Serii: Aktualni problemy rozvytku ukrainskoho suspilstva*, (1), 77–81. <https://doi.org/10.20998/2227-6890.2020.1.14>

3. Rozhkov, Yu. H. (2019). Metodyka vyokremlennia terminiv veterynarnoi medytsyny z korpusu veterynarykh tekstiv [Methodology for extracting veterinary medicine terms from a corpus of veterinary texts]. *Zakarpatski filolohichni studii*, 9(2), 33–36. [http://nbuv.gov.ua/UJRN/trphst_2019_9\(2\)_8](http://nbuv.gov.ua/UJRN/trphst_2019_9(2)_8)

4. Diak, T. P., & Hrytsiuk, Yu. I. (2024). Vykorystannia korpusnykh instrumentiv dlia vvyavlennia kluchovykh sliv u striltskykh i povstanskykh pisniakh yak zhanru folklornoho dyskursu [Using corpus tools to identify key words in riflemen and insurgent songs as a genre of folkloric discourse]. *Scientific Bulletin of UNFU*, 34(7), 60–71. <https://doi.org/10.36930/40340708>

5. Tsobenko, O. V. (2024). Diakhronichniy aspekt doslidzhennia anh lomovnoi morskoi terminolohii [Diachronic aspect of studying English maritime terminology]. *Zakarpatski filolohichni studii*, (36), 136–139. <http://zfs-journal.uzhnu.uz.ua/archive/36/25.pdf>

6. Zhulavska, O. O., & Dorohokup lia, A. O. (2020). Doslidzhennia bezosobovykh form diiesliv prohramnymy zasobamy korpusnoi linhvistyky [Studying impersonal verb forms using corpus linguistics software tools]. *Aktualni pytannia humanitarnykh nauk: Mizhvuzivskyi zbirnyk naukovykh prats molodykh vchenykh Drohobyskoho derzhavnoho pedahohichnoho universytetu imeni Ivana Franka*, 34(2), 131–139.

7. Xu, K., Feng, Y., Li, Q., et al. (2025). Survey on terminology extraction from texts. *Journal of Big Data*, 12, 29. <https://doi.org/10.1186/s40537-025-01077-x>

8. Bay, M., Bruneß, D., Herold, M., Schulze, C., Guckert, M., & Minor, M. (2020). Term extraction from medical documents using word embeddings. In *2020 6th IEEE Congress on Information Science and Technology (CiSt)* (pp. 328–333). IEEE. <https://doi.org/10.1109/CiSt49399.2021.9357263>

9. Frantzi, K., Ananiadou, S., & Mima, H. (2000). Automatic recognition of multi-word terms: The C-value/NC-value method. *International Journal on Digital Libraries*, 3, 115–130. <https://doi.org/10.1007/s007999900023>