УДК 81’32:81’367
DOI https://doi.org/10.32838/2710-4656/2022.4.2/07

Тугай О. М.
Київський університет імені Бориса Грінченка

КВАНТИТАТИВНА ВЕРИФІКАЦІЯ РАНҐОВОГО РОЗПОДІЛУ АПРОКСИМАЦІЙНОЇ ЗАЛЕЖНОСТІ В РЕЧЕННЯХ ПОСТУПКИ: КОРПУСНО-МАТЕМАТИЧНИЙ ПІДХІД

У репрезентованій статті розглянуто провідні актуальні математично-лінгвістичні методи дослідження в межах корпусного аналізу текстів, а саме автоматично скомпільованого корпусу реалізації універсальних речень поступки в художніх текстах Британського національного корпусу. В термінах корпусного підходу обчислено та отримано відповідні дистрибутивно-статистичні дані частоти появи або вживання та організації досліджуваних речень поступки. Розподіл актуалізації універсальних речень поступки у Британському національному корпусі встановлено за ранґом найбільшої / найменшої частоти вживання сполучників поступки від сполучника still до сполучника howsoever як від 1 до 31 рангу із урахуванням отриманих квантитативних показників. Провідну увагу також зосереджено на доцільності та практичності використання математичних методів та створення математичних моделей для розв’язання певних лінгвістичних задач в корпусному аналізі великих масивів текстів та метаданих. За методом Ціпфа квантитативні показники ранґового розподілу апроксимаційної залежності універсальних речень поступки, а саме частоти появи речень від рангу сполучника поступки верифіковано в наступній конфігурації: ступінь розподілу ідентифіковано як $f(x) \approx 166585$, коефіцієнт ступеня розподілу визначено як $\gamma \approx 2.65$, та, відповідно, коефіцієнт детермінації отримано як $R^2 \approx 0.83$. За методом χ^2 (хі-квадрат) визначено індикатор похибки вибіркової частоти речень універсальної поступки від середньої або апроксимаційної залежності як $\chi^2 \approx 1.15$, що сигналізує про випадкове відхилення вибіркової частоти реалізації досліджуваних речень зі сполучником поступки від апроксимаційної залежності. Отримані точні дані з урахуванням розрахунків корпусного аналізу обґрунтовано доводять релевантність залучених математичних методів для обчислення та квантитативної верифікації ранґового розподілу апроксимаційної залежності універсальних речень поступки в художніх текстах Британського національного корпусу.

Ключові слова: *універсальне речення поступки, корпусна лінгвістика, квантитативна верифікація, ранґовий розподіл, апроксимаційна залежність, Британський національний корпус.*

Постановка проблеми. Типовою ознакою сучасного мовознавства є застосування різних математичних методів та моделювання лінгвістичних систем для аналізу лінгвістичного матеріалу, вирішення статистичних задач у найрізноманітніших наукових гуманітарних дослідженнях, які відзначаються яскраво вираженим міждисциплінарним характером, що розширює межі імплементації методів одних наук в інших, як, наприклад, соціальних та гуманітарних. Причому застосування математичних підходів в лінгвістиці містить різноманітних характер, як розрахунок статистичних характеристик текстів або розробка регресійних моделей. «Моделювання – це певна універсальна процедура, яка має чітко означену кінцеву мету та суворо детермінований спосіб її досягнення». Проте методи моделювання в лінгвістиці не виникли з нічого: принцип створення моделей у лінгвістиці має «відбиток» того чи іншого магістрального підходу [2, с. 10–12].

Корпусні дослідження також виступають потужним інструментом для вивчення мови. Вони уможливлюють зробити пошук лінгвістичних даних автоматизовано; дають змогу проаналізувати мовні явища вичерпно й різноаспектно із залученням значного за обсягом матеріалу великих структурованих колекцій текстів природних мов [1, с. 17].

У нашій розвідці застосування корпусного та математичного підходів у поєднанні для аналізу реалізації універсальних речень поступки в художніх текстах Британського національного корпусу (British National Corpus – BNC) має важливе значення для чіткого розуміння характеру взаємодії та шляхів імплементації зазначених методів у лінгвістиці.

Релевантність дослідження полягає в обґрунтуванні доцільності застосування запропонованих математичних методів та моделей для квантитативної верифікації отриманих метаданих великих масивів корпусів текстів. Наочна демонстрація певних способів математичного опису зазначених мовних даних, специфіки розв'язання лінгвістичних задач та процесів за допомогою корпусного та математичного аналізу уможливить ідентифікувати особливі риси та параметри лінгвістичного моделювання для оцінки ймовірності використання певних мовних моделей в різних текстах.

Аналіз останніх досліджень і публікацій. На сьогодні корпусні та математичні дослідження в лінгвістиці мають актуальне значення, оскільки відзначаються точною обробкою метаданих великих масивів корпусів різних текстів, отриманням відповідної квантитативної верифікації певного мовного аспекту, як ранґовий розподіл словоформ у певному тексті або в мовній групі.

Як показує здійснений нами огляд літератури, корпусні дослідження загального характеру активно розглянуті та опрацьовані як в українських наукових працях – О. Ю. Андрушенко [1], В. В. Жуковської [3], так і в розвідках зарубіжний науковців – W. J. Crawford та E. Csomay [8], Yan Zhang [11]. Вузькоспрямовані специфічні питання корпусної лінгвістики висвітлено у працях Н. Бобер, Я. В. Капранова, А. Кукаріної, Т. Тронь, Т. Насаєвич [7; 4]. Феномен застосування математичних теорій в лінгвістиці та методів лінгвістичного моделювання достатньо ретельно представлено у працях О. Васильєва, І. Васильєвої, О. Чалого [2].

Методи інноваційного та математичного моделювання також неодноразово були представлені у працях українських мовознавців О. С. Колесника, Р. К. Махачашвілі, І. В. Семеніста. Так, О. С. Колесник розробив та представив універсальні моделі ірраціонального пізнання та семіозу в діахронних та крос-культурних аспектах шляхом застосування матриці математичних даних та формул із утворенням відповідних концептуальних лінгвістичних моделей для опису різних мовних аспектів [9]. Р. К. Махачашвілі та І. В. Семеніст ретельно дослідили макро- та мікроструктури глобальної інноваційної логосфери комп'ютерного буття [10].

Проте на сьогодні невирішеними залишилися питання реалізації синтаксису германських мов, задачі яких передбачають імплементацію корпусних та математичних методів, що становить певну лакуну в сучасних дослідженнях з германістики та зумовлює актуальність нашої розвідки.

Основними методами нашого дослідження слугували методи корпусного та математично-лінгвістичного аналізу.

Методика дослідження з урахуванням корпусного аналізу полягає в автоматичній вибірці універсальних речень поступки зі сполучниками поступальної дії (31 сполучник – 52973 приклади) із текстів художньої літератури сучасного Британського національного корпусу (Табл. 1).

Таблиця 1

Ранґовий розподіл сполучників універсальної поступки в художніх текстах Британського національного корпусу

Ранґовий розподіл сполучників універсальної поступки в художніх текстах Британського національного корпусу		
Ранґ сполучника поступки	**Назва сполучника поступки**	**Частота реалізації за спаданням**
1	still	16986
2	(even) though	7742
3	yet	7556
4	although	3892
5	anyway	3693
6	however	3012
7	after all	2582
8	unless	1431
9	despite	1427
10	with all	978
11	at the same time	851
12	in spite of	638
13	nevertheless	603
14	in any case	591
15	at any rate	218
16	anyhow	164
17	for all that	108
18	yet … though / though … yet	81
19	regardless of	71
20	with all that	70
21	nonetheless	66
22	in any event	48
23	after all that	40
24	yet … although / although … yet	34
25	notwithstanding	32
26	in spite of the fact that	26
27	at all events	12
28	irrespective of	8
29	despite that	6
30	nevertheless … though	6
31	howsoever	1
Загалом:		52973

Методика дослідження з урахуванням одного із фундаментальних або «класичних» методів математичної лінгвістики, а саме закону Ціпфа для ранґового розподілу слів у тексті [2, с. 10], полягає у з'ясуванні *апроксимаційної залежності логарифма частоти появи* (вживання) досліджуваних речень поступки від *логарифма ранґу* певного сполучника універсальної поступки (за функцією спадання сильної ознаки реалізації поступальної дії – від сильної актуалізації поступки (сполучник *still*) до реалізації слабкої функції поступки (сполучник *howsoever*)) – для окреслення розподілу апроксимації коливань реалізації та вживання цих речень в художніх текстах сучасного BNC – від найбільшої / найменшої кількості реалізації певного структурно-семантичного типу речення із відповідним сполучником до частоти актуалізації речень поступки зі сполучником «сильної / слабкої» функції вираження поступальної дії.

Методика аналізу матеріалу за «хі-квадрат критерієм» (χ^2 метод аналізу) полягає у верифікації *випадковості чи суттєвості відхилення або похибки* нашої вибіркової частоти від середньої для ідентифікації точності даних апроксимаційної залежності частоти сполучника універсальної поступки від його ранґу [5, с. 398-399]. У нашому дослідженні корпусний та математично-лінгвістичний підходи відіграють суттєве значення для аналізу квантитативних даних реалізації універсальних речень поступки в художніх текстах Британського національного корпусу. Корпусний та математичний інструментарій дав змогу виміряти кількісні показники актуалізації досліджуваних речень поступки та розподіл їх вживання в сучасному корпусі англійських художніх текстів.

Постановка завдання. *Об'єктом* нашої розвідки є розроблений корпус універсальних речень поступки, скомпільований із художніх текстів Британського національного корпусу. *Предметом* статті виступають дистрибутивно-статистичні характеристики вживання концесивних клауз в сучасних англійських художніх творах корпусної лінгвістики. *Метою* дослідження є застосування класичних математичних методів (закон Ціпфа та χ^2 аналіз) для обчислення та окреслення статистичних даних ранґового розподілу апроксимаційної залежності логарифма частоти появи універсальних речень поступки від логарифма ранґу вживання сполучника уведення поступальної дії в реченні.

Для реалізації мети нашої розвідки передбачаємо розв'язання таких *завдань*: 1) обґрунтувати релевантність застосування математичних методів та відповідних математичних моделей в корпусній лінгвістиці; 2) здійснити процедуру квантитативної верифікації універсальних речень поступки в художніх текстах BNC; 3) визначити розподіл актуалізації універсальних речень поступки за ранґом частоти вживання сполучника уведення поступальної дії; 4) з'ясувати та окреслити апроксимаційну залежність логарифма частоти появи речення від логарифма ранґу сполучника поступки; 5) ідентифікувати показник похибки вибіркової частоти речень універсальної поступки від середньої або апроксимаційної залежності за «хі-квадрат критерієм» – χ^2 метод – із детермінацією суттєвості чи випадковості такого відхилення. *Матеріалом* дослідження слугували виокремлені з художніх текстів Британського національного корпусу універсальні речення поступки, корпус яких склав 52973 одиниць (див. табл. 1).

Виклад основного матеріалу. У філологічних студіях корпусна лінгвістика вже зарекомендувала себе як самодостатня наукова галузь знань. Інструменти корпусу як знаряддя для керування великими масивами даних призводять до більш організованого набору слів у хаотичній різноманітності мов. Корпусний підхід можна також назвати методом дистрибутивно-статистичного аналізу, який широко використовується в методиці навчання англійської мови, і який одночасно виступає одним із методів різних галузей лінгвістичних досліджень при оцінці чи обробці даних [7, с. 176].

За О. Ю. Андрушенко, «більш репрезентативні результати даних корпусу сприяють перегляду багатьох лінгвістичних постулатів і демонструють якісно нові характеристики конкретних одиниць як однієї мови, так і багатьох мов» [1, с. 17]. На сьогодні існують монолінгвальні корпуси текстів, наприклад, *The Intelligent Web-based Corpus, British National Corpus, American National Corpus, CoRola, TS Corpus* тощо, які загалом мають обсяг близько 14 млрд. слововживань із різножанрових текстів та містять різноманітну корпусну розмітку, включаючи колігацію або узгодження, семантичне тегування, лематизацію тощо [1, с. 17; 8].

Окремі корпуси текстів мають свій діапазон слововживань. Так, як зазначають провідні українські дослідники різних аспектів корпусної лінгвістики Н. Бобер, Я. Капранов, А. Кукаріна, Т. Тронь та Т. Насалевич, Британський національний корпус містить діапазон у 100 мільйонів слів, серед яких 90% становлять письмові тексти, і лише 10% – розмовні тексти (підкорпуси), які належать

до різних жанрів кінця XX ст., а саме: зразки ділового листування, науково-популярна література, газетні статті, тексти на релігійну тематику, записи урядових промов, транскрибовані записи неофіційних передач тощо [7, с. 182]. Для кращого розуміння дистрибутивно-статистичних наукових даних у досліджуваних художніх текстах з BNC варто правильно окреслити поняття «корпусні тексти». Слідом за В. В. Жуковською, визначаємо «корпус текстів як машиночитану, збалансовану, репрезентативну колекцію спеціально позначених (анотованих) текстів, відібраних за фіксованими параметрами для досягнення певної лінгвістичної мети та досліджуваних нелінійно за принципом гіпертексту» [3, с. 58].

У нашій розвідці для застосування математично-лінгвістичних методів з метою квантитативної верифікації отриманих даних спочатку було застосовано метод корпусного аналізу великих масивів текстів. Для цього першим кроком було залучено онлайн сервіс Британського національного корпусу [6] для розробки та компіляції спеціального корпусу універсальних речень поступальної семантики зі залученням 31 сполучника поступки шляхом автоматичного пошуку за певним сполучником.

На рис. 1, 2 продемонстровано автоматичний пошук універсальних речень поступки зі сполучником *though* (за виключенням сполучуваності «as though») в художніх текстах BNC наступним шляхом [6]: 1) у розділі «Query options» головного меню вибираємо опцію «Written restrictions»; 2) у розділах «Derived text type» та «Genre» вибираємо опції «Fiction and verse» та «W:fict:drama / W:fict:poetry / W:fict:prose», відповідно; 3) у пошукове віконце «Query term» уводимо певний досліджуваний сполучник поступки; 4) виставляємо опції у відповідних віконцах «Query mode» та «Number of hits per page»; 5) натискаємо кнопку «Start query» та отримуємо результат пошуку.

Другим кроком було застосовано автоматичне обчислення квантитативних даних з визначенням відповідного рангу кожного сполучника поступки та було отримано квантитативні показники ранґового розподілу універсальних речень поступальної дії в художніх текстах BNC (див. табл. 1).

На рис. 3, 4 продемонстровано отримані обчислені результати автоматичного пошуку реалізації універсальних речень поступки зі сполучниками *though* (за виключенням сполучуваності «*as though*») та *although* в художніх текстах BNC [6]:

Рис. 1. Автоматичний пошук універсальних речень поступки зі сполучником *though* в художніх текстах BNC

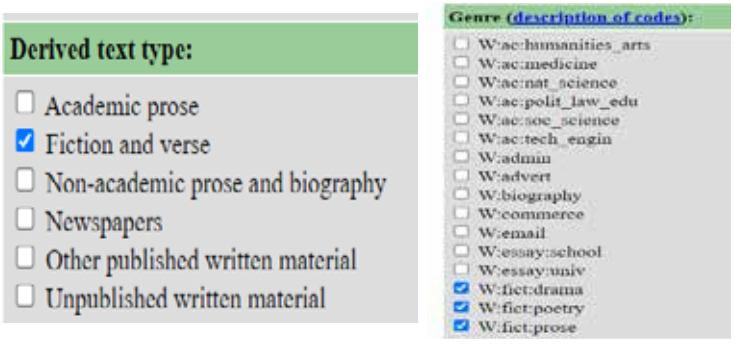

Рис. 2. Виставлення опцій текстового типу та жанру для автоматичного пошуку універсальних речень поступки в художніх текстах BNC

Your query "though" restricted to "*Derived text type: Fiction and verse* and *Genre: W:fict:drama* or *W:fict:poetry* or *W:fict:prose*" returned 10684 hits in 415 different texts (16,143,913 words [452 texts]; frequency: 661.8 instances per million words)

|< << >> >| Show Page 1 Show KWIC View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 1000 Page 1 / 11
1	A08 5	I had been preparing myself for as long as I can remember, preparing myself (though I did not always realize it) from the day that I was born, preparing myself, wrote Harriet (typed Goldberg), but always aware of the dangers of beginning too soon.
2	A08 13	That is why, wrote Harriet, I have been preparing myself for that moment for a long time, that is why I have cleared the decks and prepared the ground , because unless the decks are cleared and the g round prepared there is little hope of succeeding in what one has planned to do, little hope of achieving anything of lasting value, though lasting is a relative term and so is value and whatever it is one has planned to do is certain to be altered in the process, which does not of course mean, he wrote, that one can start anywhere at any time.
3	A08 15	It is just because whatever one has planned is bound to change as one proceeds that it is fatal to start too soon or too late. though it may be no less fatal, he wrote (and Goldberg typed), to start at the right time, for then there is no excuse, no excuse whatsoever.
4	A08 29	Everything possible must be done, he wrote, and yet it must be as though nothing had been done.
5	A08 41	It was neither pleasant nor unpleasant, though the endless pacing, he wrote, the endless getting up in the middle of the night when the ice clung to the windowpanes and the taps were frozen, that was more unpleasant than pleasant, but it was not that, he wrote, these things - will not change, my bladder will not improve and next winter the ice will still cling to the panes and the taps will still freeze, but I will not notice them .
6	A08 44	But I do not mean to suggest either, he wrote, that it was all waiting and no doing, all sitting and no action, for though it was impossible to tell when the beginning would come, indeed, he wrote, there could not have been a real beginning if it had been possible to tell, for if it had been possible to tell that would have meant that there had already been a beginning, so, wrote Harriet (typed Goldberg), occasionally things were done, work was begun, though it was soon abandoned, it added up to nothing, it only showed me that I had been mistaken in thinking that I had indeed started.
7	A08 44	But I do not mean to suggest either, he wrote, that it was all waiting and no doing, all sitting and no action, for though it was impossible to tell when the beginning would come, indeed, he wrote, there could not have been a real beginning if it had been possible to tell, for if it had been possible to tell that would have meant that there had already been a beginning, so, wrote Harriet (typed Goldberg), occasionally things were done, work was begun, though it was soon abandoned, it added up to nothing, it only showed me that I had been mistaken in thinking that I had indeed started.
8	A08 56	That is a fact, he wrote (and Goldberg typed), one of the few facts I can swear to, though I find it impossible to explain.
9	A08 57	And it has to be said, he wrote, that its opposite, a feeling of elation, equally physical, equally extra-physical, has also been a constant feature of my life, manifesting itself regularly though impossible to predict, a swelling in the chest this time, the chest and perhaps the throat, a feeling of the heart leaping and the blood pumping, it came when I first took up a brush and made a mark on paper, it came when I picked up the first readymade and felt it transformed by that very action, it came when Madge rang to say she could not go on, when Annie wrote to say she was not coming back, when the idea of the glass first popped into my head.
10	A08 59	Taken in by the image of yourself they present you with, wrote Harriet, instead of waiting in patience for the beginning, instead of waiting and then beginning, though beginning, having begun, he wrote, is not everything, is far from everything.
11	A08 61	Though it may well be, he wrote, that one actually achieves more working with the wrong plans and in the wrong spirit, with the wrong tools and the wrong principles, on the wrong surface and with the wrong conception, it may well be, he wrote (and Goldberg typed), that one achieves more than working with the right plans and in the right spirit, with the right tools and the right principles, on the right surface and with the right conception, though right and wrong and more and less are relative concepts and what seems right at one moment to one person may seem wrong at the same moment to another Person or at another moment to the same person, and what seems more to one person at one moment may seem less to another person at the same moment or at another moment to the same person, right, wrong, more, less, relative concepts , scribbled Goldberg, in the margin, panting slightly as he bent over his old Olivetti Portable, there is only the beginning, wrote Harriet, or rather, there is only having begun, beginning , scribbled Goldberg, aware now of the black stains on his hands left by the felt-tip pen, having begun , there is only the feeling in the pit of the stomach or the feeling in the chest, wrote Harriet, the feeling of sickness or the feeling of elation, those are not relative, he wrote, those are absolute.

Рис. 3. Загальний результат автоматичного пошуку універсальних речень поступки зі сполучником *though* (без урахування сполучуваності «*as though*») в художніх текстах BNC

Your query "although" restricted to "*Derived text type: Fiction and verse* and *Genre: W:fict:drama* or *W:fict:poetry* or *W:fict:prose*" returned 3892 hits in 363 different texts (16,143,913 words [452 texts]; frequency: 241.08 instances per million words)

|< << >> >| Show Page 1 Show KWIC View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 1000 Page 1 / 4
1	A0D 117	If we don't, although now you get the estate you can't, of course, inherit the title.'
2	A0D 417	On board the steamer the two of them were talking about what would happen to the title if Lord Woodleigh was to die before they had any children, and Lord Woodleigh said — Sven Hjerson's ears heard it — although now you get the estate .
3	A0D 1272	As did the House Manager who roamed throughout performances in the foyer or the staircases, the bars of Stalls, Dress Circle and Upper Circle, keeping an eye on programme girls (most of them certainly mature) who, in their black dresses and little aprons, tidyed, sold programmes and in the intervals brought trays of tea and biscuits (coffee in the evenings), while in the orchestra pit the band (tuxedoed, although who knew whether their trousers matched) played pleasing music.
4	A0D 1569	But, although it was something to tell the others at school, secretly I thought he was important enough already.
5	A0D 1707	Although , at that moment I could have done with a little less myself.
6	A0D 2598	Otherwise you could have seen the garden, although there's not much to look at at this time of the year.'
7	A0D 2919	'More of a horticultural poison,' said the inspector, 'although not intended as such.'
8	A0F 579	Although I can now see the inherent sense in ghosting a book to a table, then the point was just beyond me.
9	A0F 624	Although Jeff making me laugh at myself was the beginning of the end of my depression, it wasn't enough to persuade me to stay.
10	A0F 874	However, although unemployment was starting to come down, especially in the south-east where I was living, it was still generally pretty high.
11	A0F 1284	Although I hadn't seen her in over thirty years, there was one person in the world who might help me: my sister.
12	A0F 1339	Although their house was still some distance away, I decided to finish the journey on foot.
13	A0F 2127	Both Jenny and Kathleen were up to various things and I'd been having a go myself, but although my name was appearing on more and more waiting lists, nothing substantial seemed to be happening.
14	A0L 15	She'd taken him literally, although now she questioned whether that had been wise.
15	A0N 1474	You will manage — oh yes — although you have no servants to feed your animals or cook your meat for you — you will manage , although you are a year married with a wee child learning to walk, and no wet-nurse to mind it for you while you milk the beasts — and no young husband to thatch your house above your head.
16	A0N 1474	You will manage — oh yes — although you have no servants to feed your animals or cook your meat for you — you will manage , although you are a year married with a wee child learning to walk, and no wet-nurse to mind it for you while you milk the beasts — and no young husband to thatch your house above your head.
17	A0N 1529	The lawyers at the government mill at Kinlochanoch were good friends of Cameron's, and might try to bring out their neighbours, although they themselves were still felt to be incomers.
18	A0R 46	The brooch and the bag would make it clear she had originality although, taken in the context of the rest of the outfit, not too much.
19	A0R 120	She made no effort to turn over, although the thought about it, imagined how it might be to lean on one elbow, to twist her body in a single movement.
20	A0R 224	But unfortunately it was misspelt far more often than peas or spring greens or even aubergine, although that, too, was a difficult word.
21	A0R 238	When she arrived at the bus station she saw on the wall behind her bold, splashy writing in foreign characters, Arabic maybe or Urdu, and small, disordered scribbles around the glass faces of the timetables, which, although an imitation, caused Rita no real pain.
22	A0R 250	The words on the pavement were common currency nowadays, although Rita's mother would have found them deeply offensive.

Рис. 4. Точний результат автоматичного пошуку універсальних речень поступки зі сполучником *although* в художніх текстах BNC

Відповідно, кількісний показник частоти реалізації сполучника або прислівника *still* із позначкою 16986 одиниць реалізації універсальних речень поступки у BNC зумовив отримання ним логарифма першого рангу, що сигналізує про сильний ступінь вираження функції поступки із цим сполучником, тоді як квантитативний показник частоти актуалізації сполучника *howsoever* із позначкою як 1 одиниця розподілення універсальних речень поступки у BNC став тригером отримання ним останнього логарифма 31 рангу серед всіх інших сполучників поступки, що сигналізує про слабкий ступінь реалізації функції поступки із зазначеним конектором (див. табл. 1).

Для з'ясування квантитативної верифікації апроксимаційної залежності логарифма частоти появи універсального речення поступки від логарифма рангу сполучника уведення поступальної

дії за функцією «сильної / слабкої» риси реалізації поступальної дії було залучено універсальний математичний прийом – «класичний» метод, відомий як закон Ціпфа, який пов'язує частоту слова F із його ранґом n. Це співвідношення відповідності має такий вигляд:

$$F = \frac{A}{n^{\gamma}}, \quad (1)$$

де γ є параметром або коефіцієнтом ступеня розподілу і в багатьох випадках має близьке до одиниці значення;

$$F = \frac{A}{(n + n_0)^{\gamma}}, \quad (2)$$

із поправкою Мандельброта – уведенням додаткового параметру розподілу n_0 [2, с. 14-15].

Також закон Ціпфа може бути представлений лінійною залежністю між натуральним логарифмом частоти появи слова $ln(F)$ та натуральним логарифмом ранґу слова $ln(n)$ у такому вигляді як [2, с. 15]:

$$ln(F) = ln(A) - \gamma\, ln(n). \quad (3)$$

У нашому дослідженні залучаємо формулу закону Ціпфа із поправкою Мандельброта для обчислення постійної величини залежності частоти появи універсальних речень поступки від ранґу певного сполучника (див. Табл. 2), що

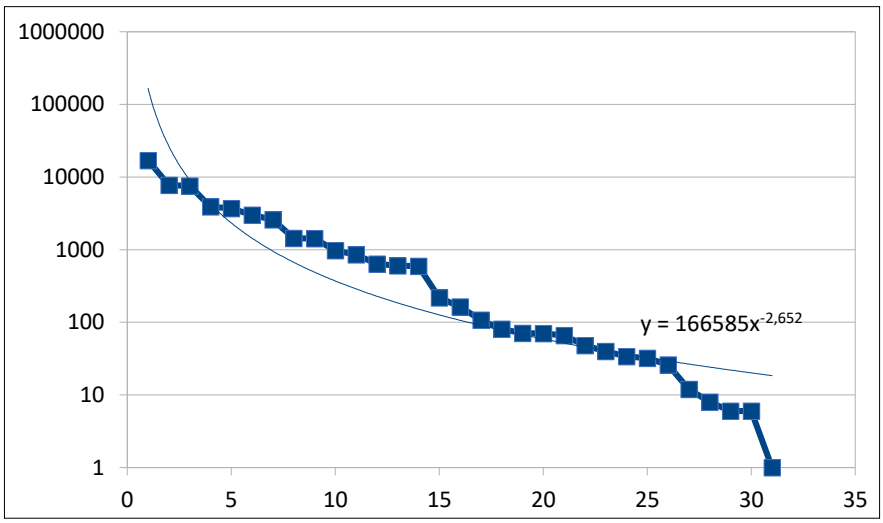

Рис. 5. Логарифмічна шкала залежності кількості вживань речень поступки від ранґу сполучника поступки в художніх текстах BNC

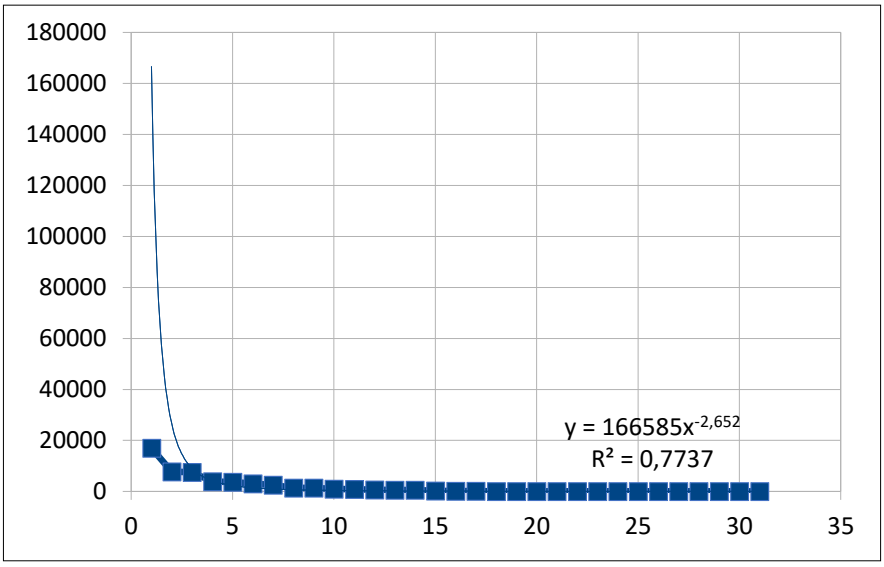

Рис. 6. Лінійно-вертикальна шкала залежності кількості вживань речень поступки від ранґу сполучника поступки в художніх текстах BNC

зображено на графіках рис. 5, 6: на горизонтальній осі відкладаємо *логарифм рангу* вживання сполучника поступки від сильної до слабкої функції вираження поступальної дії в реченні; на вертикальній осі зображуємо *логарифм частоти* реалізації універсальних речень поступки в художніх текстах BNC.

Зокрема, загальний обсяг скомпільованого нами корпусу універсальних речень поступки в художніх текстах Британського національного корпусу становить 52973 речень, кількість досліджуваних сполучників поступки дорівнює 31 одиниці. І, відповідно, обчислені значення складають: $A \approx 166585$ та $\gamma \approx 2.65$ для параметрів розподілу, що входять у закон Ціпфа. При цьому коефіцієнт детермінації R^2 складає 0.83 – що за законом Ціпфа (де параметр розподілу має бути < 2) є добрим результатом та хорошим релевантним показником ступеня розподілу частоти вживання від рангу сполучника універсальної поступки.

Частотний розподіл відповідності вживання сполучника поступки в художніх текстах BNC представлено на графіках у рис. 5 та рис. 6.

Для перевірки точності даних апроксимаційної залежності логарифма частоти від логарифма рангу сполучника поступки, а також як інструмент для детермінації випадковості чи суттєвості відхилення вибіркової частоти від середньої або апроксимації залежності було також залучено метод обчислення за «хі-квадрат критерієм» – χ^2, де «хі-квадрат» дорівнює сумі квадратів відхилень від апроксимації залежності, поділеної на середню частоту із залученням наступної формули [5, с. 398-399]:

$$\chi^2 = \frac{\sum\left(x_i - x\right)^2}{x} \qquad (4)$$

У нашому дослідженні підставляємо під вищезазначену формулу χ^2 показники кількості та рангу в наступній конфігурації:

$$\chi^2 = \frac{\sum\left(N_i - f_i\right)^2}{f_i}. \qquad (5)$$

Або в лінійній конфігурації:

$$\chi^2 = \sum\left(N_i / f_i - 1\right)^2 \qquad (6)$$

Як результат, отримуємо наступні показники: $\Sigma \approx 36.95$; $\chi^2 \approx 36.95$ – як ненормоване значення похибки апроксимації залежності, а при застосуванні операції ділення Σ (36.95) на кількість рангів сполучника універсальної поступки (31) отримуємо *нормоване значення похибки* апроксимаційної залежності як $\chi^2 \approx 1.15$. Відповідно, значення χ^2 як 1.15 характеризуємо як *випадкове відхилення* вибіркової частоти вживання сполучника універсальної поступки від апроксимаційної залежності логарифма частоти від логарифма рангу. Це доводить доречність та точність отриманих даних з нашої вибірки щодо вживання універсальних речень поступки в художніх текстах Британського національного корпусу.

Висновки і пропозиції. Отже, у нашій науковій розвідці обґрунтовано та доведено доцільність використання математично-лінгвістичних методів, таких як закон Ціпфа та метод «хі-квадрату» для обчислення дистрибутивно-статистичних даних з автоматично скомпільованого корпусу універсальних речень поступки в художніх текстах Британського національного корпусу. З урахуванням корпусного аналізу розподіл реалізації універсальних речень поступки у BNC визначено в термінах найбільшої / найменшої частоти вживання сполучників поступки від сполучника *still* до сполучника *howsoever* за рангом частоти їхнього вживання від 1 до 31. З'ясовано, що розподіл апроксимаційної залежності логарифма частоти появи універсальних речень поступки в художніх текстах BNC від логарифма рангу кожного сполучника має точні дані квантитативної верифікації, а саме: ступінь розподілу $f(x) \approx 166585$, коефіцієнт ступеня розподілу $\gamma \approx 2.65$, коефіцієнт детермінації $R^2 \approx 0.83$ при похибці апроксимаційної залежності $\chi^2 \approx 1.15$, що є показником випадкового відхилення вибіркової частоти реалізації речень зі сполучником поступки від апроксимаційної залежності. Це доводить відповідність вживання частоти або кількості речень універсальної поступки із рангом сполучника поступки в художніх текстах BNC, а також доцільність та точність залучених математичних розрахунків. Наступні розвідки у царині корпусної лінгвістики вбачаємо в залученні математично-статистичного аналізу великих масивів корпусних текстів речень поступальної дії різної семантики у германських мовах в синхронній та діахронній площинах.

Список літератури:

1. Андрушенко О. Ю. Інформаційно-структурні перетворення адитивного адверба *EVEN* (на матеріалі пам'яток і текстів корпусів англійської мови XII–XVII ст.). *Вісник КНЛУ. Серія Філологія.* Том 24. № 1. 2021. С. 16–32.

2. Васильєв О., Чалий О., Васильєва І. Математичні методи та моделі в лінгвістиці. *Україна модерна.* № 27. 2019. С. 9–28.

3. Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир : Вид-во ЖДУ ім. І. Франка, 2013. 142 с.

4. Капранов Я. В. Презентація наукових результатів квантитативної нестратичної верифікації ступенів споріднення афразійської, індоєвропейської і картвельської мовних сімей. *Вісник КНЛУ. Серія Філологія.* Том 23, № 2. 2020. С. 58–71.

5. Кочерган М. П. Загальне мовознавство: підручник. Видання 2-ге, виправлене і доповнене. Київ : Видавничий центр «Академія», 2006. 464 с.

6. BNC Web at Lancaster University. URL: http://bncweb.lancs.ac.uk/ (Last accessed: 23.08.2022).

7. Bober N., Kapranov Y., Kukarina A., Tron T., Nasalevych T. British National Corpus in English language teaching of university students. I*nternational Journal of Learning, Teaching and Educational Research.* June 2021. Vol. 20, No. 6. P. 174–193.

8. Crawford W. J., Csomay E. Doing corpus linguistics. New York – London: Routledge, Taylor & Francis Group, 2016. 178 p.

9. Kolesnyk O. The cognitive premises of myth-oriented semiosis. *Cognitive Studies | Études cognitives,* 2019 (19).

10. Makhachashvili R. K., Semenist I. V. Phenomenological paradigm of digital innovative logosphere modelling (based on innovations of the Chinese language). *New Philology,* (85), 2022. P. 173–180.

11. Yan Zhang. Adversative and Concessive Conjunctions in EFL Writing: Corpus-based Description and Rhetorical Structure Analysis. 1st ed. Singapore, Shanghai : Springer Nature Singapore Pte Ltd. & Shanghai Jiao Tong University Press, 2021. 234 p.

Tuhai O. M. QUANTITATIVE VERIFICATION OF RANK DISTRIBUTION OF APPROXIMATION DEPENDENCE IN CONCESSIVE SENTENCES: CORPUS-MATHEMATICAL APPROACH

The presented article deals with the leading topical mathematical and linguistic methods of research within the framework of corpus analysis of texts, namely the automatically compiled corpus of universal concessive sentences realization in the written fiction texts of the British National Corpus. In terms of the corpus approach, the corresponding distributional and statistical data of the frequency of occurrence or usage and organization of the studied concessive sentences are calculated and obtained. The distribution of universal concessive sentences actualization in the British National Corpus is determined by the rank of the highest / lowest frequency of concessive conjunctions usage from conjunction still to conjunction howsoever as from the 1st to the 31st rank, taking into account the obtained quantitative indicators. Leading attention is also focused on the expediency and practicability of mathematical methods usage and creating mathematical models for solving certain linguistic problems in the corpus analysis of large arrays of texts and metadata. According to the Zipf's method, the quantitative indicators of the rank distribution of the approximation dependence of universal clauses of concession, namely, frequency of concessive sentences' occurrence from the rank of the conjunction of concession, are verified in the following configuration: the degree of distribution is identified as $f(x) \approx 166585$*, the coefficient of the degree of distribution is determined as* $\gamma \approx 2.65$*, and accordingly, the coefficient of determination is obtained as* $R^2 \approx 0.83$*. According to the* χ^2 *(chi-square) method, the indicator of the error of the sampling frequency of universal concessive sentences from the average or approximation dependence is determined as* $\chi^2 \approx 1.15$*, which signals about a random deviation of the sampling frequency of the studied sentences with concessive conjunction implementation from the approximation dependence. The obtained accurate data, taking into account the calculations of the corpus analysis, reasonably prove the relevance of the involved mathematical methods for the calculation and quantitative verification of the rank distribution of the approximation dependence of the universal sentences of concession in the written fiction texts of the British National Corpus.*

Key words: *universal concessive sentence, corpus linguistics, quantitative verification, rank distribution, approximation dependence, British National Corpus.*