

РОЗДІЛ 4

СТРУКТУРНА, ПРИКЛАДНА ТА МАТЕМАТИЧНА ЛІНГВІСТИКА

УДК 81.33

DOI <https://doi.org/10.32782/tps2663-4880/2022.24.2.29>

КОРПУС СУЧАСНОЇ АМЕРИКАНСЬКОЇ АНГЛІЙСЬКОЇ МОВИ ЯК ІНСТРУМЕНТ ЛІНГВІСТИЧНИХ ДОСЛІДЖЕНЬ

CORPUS OF CONTEMPORARY AMERICAN ENGLISH AS A TOOL OF LINGUISTIC RESEARCH

Бігдай М.О.,

*orcid.org/0000-0001-6507-6100**кандидат філологічних наук,**асистент кафедри прикладної лінгвістики**Національного університету «Львівська політехніка»*

Мета статті – описати основні функції Корпусу сучасної американської англійської мови (Corpus of Contemporary American English, COCA) для лінгвістичних досліджень на прикладі прикметників ‘quiet’ і ‘peaceful’.

Незважаючи на те, що на сьогоднішній день існує багато корпусних студій у різних країн світу, проте в українському мовознавстві Корпус сучасної американської англійської мови як інструмент лінгвістичних досліджень не був достатньо досліджений. Недостатність опису цієї проблематики зумовлює актуальність нашої розвідки.

Окрім цього, актуальність теми зумовлена розвитком корпусної лінгвістики, різноманіттям корпусів, що постійно з'являються і визначається необхідністю проведення досліджень на основі американської англійської мови на великому масиві даних.

Корпус сучасної американської англійської мови містить близько 1 млрд слів з 1990 до 2019 років. Особливістю цього корпусу є рівномірний жанровий розподіл між такими жанрами, як: розмовні тексти, художня література, популярні журнали, газети, академічні журнали, телебачення/субтитри фільмів, блоги, веб-сторінки. У кожному жанрі приблизно 120-130 мільйонів слів.

Проаналізувавши словосполучення ‘quiet + іменник’ та ‘peaceful + той самий іменник’, доходимо до висновку, що перші трапляються в корпусі COCA набагато частіше (загальна частота 3.50 на противагу 0.29).

Корпус сучасної американської англійської мови може слугувати матеріалом для проведення різноманітних лексикологічних досліджень для американської англійської мови, а саме: вивчення контекстів, частотності вживання слів, словосполучень, граматичних конструкцій, багатозначності, омонімії, синонімії, антонімії, історичних змін в словниковому складі мови. Перспективу подальших розвідок вбачаємо у дослідженні інших частин мови на основі даних цього корпусу.

Ключові слова: корпусна лінгвістика, корпус, статичний корпус, динамічний корпус, американська англійська мова.

The purpose of the article is to describe the main functions of the Corpus of Contemporary American English (COCA) for linguistic research using the adjectives ‘quiet’ and ‘peaceful’ as an example.

Despite the fact that today there are many corpus studies in different countries of the world, the Corpus of Contemporary American English as a tool of linguistic research has not been sufficiently studied in Ukrainian linguistics. The lack of the description of this issue determines the relevance of our investigation.

In addition, the topicality of the theme is determined by the development of corpus linguistics, the variety of corpora that constantly appear and is determined by the need to conduct research based on American English on a large array of data.

The Corpus of Contemporary American English contains about 1 billion words from 1990 to 2019. A peculiarity of this corpus is an even genre distribution among such genres as: spoken, fiction, popular magazines, newspapers, academic journals, television/film subtitles, blogs, web pages. Each genre has approximately 120-130 million words.

After analyzing the collocations ‘quiet + noun’ and ‘peaceful + same noun’, we came to the conclusion that the former occurs much more often in the COCA corpus (total frequency 3.50 against 0.29).

The Corpus of Contemporary American English can serve as material for conducting various lexicological studies for the American English language, namely: the study of contexts, frequency of use of words, collocations, grammatical constructions, polysemy, homonymy, synonymy, antonymy, historical changes in the vocabulary of the language. We see the prospect of further investigations in the study of other parts of the language based on the data of this corpus.

Key words: corpus linguistics, corpus, static corpus, dynamic corpus, American English.

Постановка проблеми. Корпусна лінгвістика є одним із пріоритетних напрямків сучасної лінгвістичної науки. Розвиток корпусної лінгвістики тісно пов'язаний із розвитком комп'ютерних технологій та інформатизацією. Корпусні студії дозволяють максимально об'єктивно вивчати різноманітні мовні явища, оскільки відбуваються на матеріалах текстів природної мови в умовах реального функціонування.

Актуальність теми зумовлена розвитком корпусної лінгвістики і визначається необхідністю проведення досліджень на основі американської англійської мови на великому масиві даних.

Аналіз останніх досліджень і публікацій. На сьогоднішній час здійснюється велика кількість досліджень у галузі корпусної лінгвістики. Корпусні студії проводяться дослідниками багатьох країн світу (О. М. Демська-Кульчицька [1, 2, 3], В. В. Жуковська [4], Є. А. Карпіловська [5], М. О. Шведова [6], В. А. Широков [7], D. Biber [8], W. Francis [9], G. Leech [10], J. Sinclair [11] та багато інших).

Термін «корпусна лінгвістика» вперше згаданий у 1983 році у збірнику наукових праць «Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research» [4, с. 8].

Завданням корпусної лінгвістики є розробити теоретичні засади і практичні прийоми побудови, машинного опрацювання та експлуатації лінгвальних даних у формі корпусу текстів [1, с. 45].

Корпусна лінгвістика має на меті розв'язання таких проблем:

- спосіб представлення та збереження мовленнєвих одиниць;
- вимоги до корпусу текстів із боку укладачів і користувачів;
- специфіка програмного забезпечення корпусів;
- принципи добору параметрів проблемної галузі;
- методи структурування корпусу;
- транскрипція для текстів усного мовлення;
- мультимедійна підтримка для корпусів усного мовлення;
- розробка систем пошуку у корпусі;
- кодування дескрипцій одиниць збереження тощо [12, с. 669].

За словами О. О. Селіванової «головним поняттям корпусної лінгвістики є корпус мовленнєвої реалізації мови, що кваліфікується як сформована за певними вимогами вибірка мовленнєвого матеріалу, яку можна використовувати для опису й дослідження мови як системи» [12, с. 668].

Під корпусом розуміють корпус «машиночитане, стандартно організоване зібрання репрезентативних для певної мови, діалекту або іншої підмножин(и) мов(и) писемних або усних текстів, призначених для лінгвістичного аналізу й опису, відібраних і впорядкованих згідно з експліцитними екстра- та інтралінгвальними критеріями» [2, с. 53].

До основних ознак корпусу відносять репрезентативність, автентичність, відібраність, збалансованість, машиночитаність:

- Репрезентативність полягає в тому, наскільки в корпусі відображені всі властивості певної предметної галузі. Предметна галузь містить феномени, що підлягають лінгвістичному описові.

- Автентичність передбачає відбір текстів усного або писемного мовлення, записаних у процесі реальної комунікації.

- Відібраність передбачає застосування конкретних правил відбору даних, що відповідають певній стратегії побудови корпусу в залежності від його типу і мети створення.

- Збалансованість полягає у пропорційній кількості текстів у корпусах.

- Машиночитаність полягає у тому, що сучасний текстовий корпус обов'язково має бути в електронній формі, повинен мати кодування первинних корпусних даних та лінгвістичну анотацію [3, с. 102-103].

Постановка завдання. Мета статті – описати основні функції Корпусу сучасної американської англійської мови (Corpus of Contemporary American English, COCA) для лінгвістичних досліджень на прикладі прикметників 'quiet' і 'peaceful'.

Виклад основного матеріалу. На сьогоднішньому етапі свого існування комп'ютерні корпуси застосовують як в лінгвістичних, так і в нелінгвістичних галузях:

- Лінгвісти-теоретики – у якості експериментальної бази для перевірки гіпотез і доведення своїх теорій.

- Прикладні лінгвісти (викладачі, перекладачі) для навчання іноземній мові і розв'язання своїх професійних завдань.

- Комп'ютерні лінгвісти для створення комп'ютерних моделей мови на основі виокремлених статистичних й лінгвістичних закономірностей.

- У соціолінгвістичних дослідженнях для вивчення мовного розмаїття (соціолектів).

- Інші фахівці з мови (літературознавці, редактори) також звертаються до даних корпусу.

- У літературознавстві, перекладознавстві, судовій лінгвістиці для аналізу дискурсу.

- Історики, соціологи для дослідження таких параметри текстів, як період, автор або жанр.

- Для розробки й налаштування автоматизованих систем з машинного перекладу, розпізнавання мовлення, інформаційного пошуку [4, с. 105-106].

Корпус сучасної американської англійської мови містить близько 1 млрд слів з 1990 до 2019 років. Особливістю цього корпусу є рівномірний жанровий розподіл між такими жанрами, як: розмовні тексти, художня література, популярні журнали, газети, академічні журнали, телебачення/субтитри фільмів, блоги, веб-сторінки. У кожному жанрі приблизно 120-130 мільйонів слів.

Статичний корпус засвідчує стан мови на певному синхронному зрізі, а динамічний корпус постійно оновлюється, що дозволяє відстежувати зміни у мові [4, с. 62]. Корпус сучасної американської англійської мови належить саме до динамічного типу.

Розглянемо можливості корпусу COCA на прикладі прикметників 'quiet' та 'peaceful' та словосполучень з ними.

Існує шість основних способів пошуку в корпусі [13]:

- По-перше, можна шукати конкретні фрази та за певним шаблоном. Наприклад: *got VERB-ed*, *BUY * ADJ NOUN*, різноманітні фразеологічні дієслова.

- По-друге, можна переглядати список із 60 000 найпопулярніших слів у корпусі, включаючи пошук за формою слова, частиною мови і навіть за значенням чи вимовою. Це особливо корисно для тих, хто вивчає мову та для викладачів.

- По-третє, для тих, хто цікавиться академічною англійською, можна переглядати детальні записи для кожного з 3000 слів Академічного списку слів (Academic Vocabulary List), укладеного у 2013 році.

- По-четверте, можна здійснювати пошук за окремим словом і переглядати словосполучення, кластери та споріднені слова для кожного з цих слів.

Окрім цього, як бачимо на Рис. 1, у корпусі COCA можна отримати велику кількість інформації про слово, включаючи інформацію про частотність, визначення, переклад, покликання на аудіо, зображення та відео, а також синоніми, пов'язані словоформи, словосполучення, кластери та конкорданси / слова в контексті (KWIC).

Рис. 1. Сторінка з детальною інформацією про слово 'quiet' у корпусі COCA

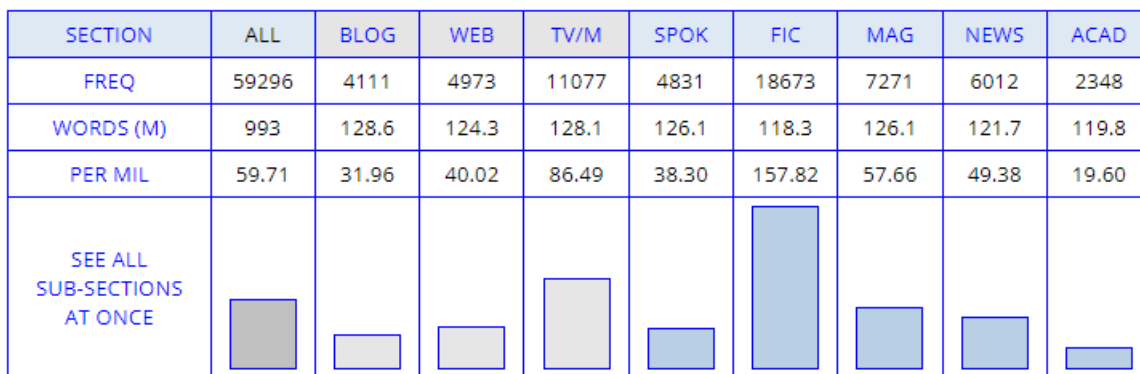


Рис. 2. Частотний розподіл слова ‘quiet’ у різних жанрах у корпусі СОСА

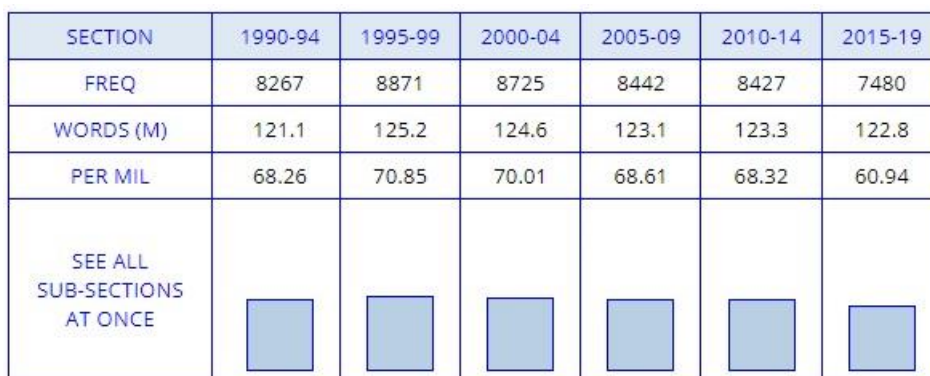


Рис. 3. Частотний розподіл слова ‘quiet’ у різних роках у корпусі СОСА

- По-п’яте, можна вводити цілі тексти, а потім використовувати дані корпусу для отримання детальної інформації про слова та фрази в тексті.

- Можна шукати та досліджувати випадкові слова, а також переглянути випадково вибрані «Слова дня».

У корпусі СОСА можна також переглядати частотність слів, фраз і граматичних конструкцій за різними жанрами, як на Рис. 2, зокрема досліджувати, чи вони належать до неформальних жанрів (наприклад, субтитри), більш офіційних жанрів (наприклад, академічні статті) або десь посередині (наприклад, журнали та газети).

Окрім цього, у корпусі СОСА можна бачити історичні дані за останні 30 років, як на Рис. 3, а також дані детально для кожного діапазону років.

На Рис. 4. бачимо у контексті словосполучення ‘quiet time’ із зазначенням жанру, назви та року випуску джерела.

Однією з найцікавіших функцій у корпусі СОСА є порівняння двох словосполучень, наприклад, коли потрібно дослідити слова синоніми.

На Рис. 5 зображено таке порівняння для словосполучень ‘quiet + іменник’ та ‘peaceful + іменник’. Обидві таблиці посортовані за частотою появи кожного іменника з досліджуваними прикметниками ‘quiet’ і ‘peaceful’. Частота словосполучення ‘quiet time’ – 620, а ‘peaceful time’ – 50. Співвідношення частот цих словосполучень (W1/W2) – 12.4. На другому місці за частотою є словосполучення ‘quiet place’ (508), ‘peaceful place’ (208) та їхнє співвідношення (W1/W2) – 2.4. І третє словосполучення ‘quiet life’ з частотою 323 та ‘peaceful life’ з частотою 165 та співвідношенням (W1/W2) – 2.0. Отже, словосполучення ‘quiet + іменник’ (загальна частота 3.50) трапляються в корпусі СОСА набагато частіше, ніж ‘peaceful + той самий іменник’ (загальна частота 0.29).

Висновки. Отже, Корпус сучасної американської англійської мови – це унікальний, великий, динамічний корпус, який постійно оновлюється і має рівномірний жанровий розподіл текстів. Цей корпус може слугувати матеріалом для проведення різноманітних лексикологічних досліджень для американської англійської мови, а саме: вивчення контекстів, частотності вжи-

1	2012	WEB	marieforleo.com	📍 📱 🔍	long walks, meditating, driving, whatever it is for each person, that quiet time to contemplate and let the desires bubble up, and we know it's
2	2012	WEB	lyricsfreak.com	📍 📱 🔍	15:17:33 Tour the lights, then some quiet time # To be honest, the plans are somewhat dependent on the weather.
3	2012	WEB	askville.amazon.com	📍 📱 🔍	and take pix of the Christmas lights or stay inside and just enjoy some quiet time in front of the fireplace light. There is definitely something about the
4	2012	WEB	therestofeverest.com	📍 📱 🔍	coming up, behind them and was bright red. It has been the only quiet time that I have had in recent days. I hope that you are experiencing
5	2012	WEB	psychologytoday.com	📍 📱 🔍	to be more open when I was with them. Had I not staked out quiet time , I would have had to stay partly shut down all the time in
6	2012	WEB	huffingtonpost.com	📍 📱 🔍	up 7:30 to 8:00 am... don't talk to me... it is my quiet time ... coffee and sit on our balcony, if there are jobs for me
7	2012	BLOG	growingsideways.net	📍 📱 🔍	tracks, rolling in boredom, and peaking up the stairs to ask me when quiet time is going to be over. # Our routine of late has been that
8	2012	BLOG	growingsideways.net	📍 📱 🔍	a way to create a carrot with which to bait jay into good behavior during quiet time . Plus, as someone who watched a lot of afternoon cartoons growing
9	2012	BLOG	patheos.com	📍 📱 🔍	group recitations -- too fast/too slow/too perfunctory. Get annoyed before evening mass when my quiet time is disturbed by some murmurously mumi
10	2012	BLOG	healthy-spirit.org	📍 📱 🔍	distractions, no television or books, no computers, no other people, just quiet time . Not the time where you sit and take a break and remember that
11	2012	BLOG	healthy-spirit.org	📍 📱 🔍	to call someone or that you have a busy day tomorrow. I mean real quiet time , where your head is empty and inspiration can come in. I know
12	2012	BLOG	healthy-spirit.org	📍 📱 🔍	continue. Every day, just for a short while. # In your own quiet time , ask yourself what you would like to change, what you are passionate
13	2012	BLOG	recipegirl.com	📍 📱 🔍	wedding, when catching up with dear friends, or when I am having sweet quiet time with my Bible and some worship music. (does that count for how
14	2012	BLOG	saportareport.com	📍 📱 🔍	, Dan Matthews was reading in his office when the phone rang, disturbing his quiet time . The current rector of downtown Atlanta's St. Luke's Episcopal
15	2012	BLOG	...omey.wordpress.com	📍 📱 🔍	laundry room to get the highly revered milk ready, already in anticipation of the quiet time alone with my boy. There was probably some sort of goofy g

Рис. 4. Словосполучення 'quiet time' у контексті у різних стилях у корпусі СОСА

WORD 1 (W1): QUIET (3.50)					WORD 2 (W2): PEACEFUL (0.29)						
	WORD	W1	W2	W1/W2	SCORE		W2	W1	W2/W1	SCORE	
1	TIME	620	50	12.4	3.5	1	RESOLUTION	379	1	379.0	1,324.6
2	PLACE	508	208	2.4	0.7	2	SOLUTION	373	0	746.0	2,607.4
3	LIFE	323	165	2.0	0.6	3	PROTEST	285	9	31.7	110.7
4	VOICE	312	2	156.0	44.6	4	COEXISTENCE	261	0	522.0	1,824.4
5	MAN	273	37	7.4	2.1	5	MEANS	252	1	252.0	880.8
6	ROOM	269	8	33.6	9.6	6	TRANSITION	221	2	110.5	385.2
7	STREET	209	9	23.2	6.6	7	SETTLEMENT	216	8	27.0	94.4
8	MOMENT	206	27	7.5	2.2	8	PURPOSES	213	0	426.0	1,488.9
9	MOMENTS	195	16	12.2	3.5	9	PLACE	208	508	0.4	1.4
10	NIGHT	193	34	5.7	1.6	10	PROTESTS	200	3	66.7	233.0
11	DAY	182	24	7.5	2.2	11	WORLD	185	29	6.4	22.3
12	CORNER	172	9	19.1	5.5	12	LIFE	165	323	0.5	1.8
13	WAY	172	157	1.1	0.3	13	WAY	157	172	0.9	3.2
14	NEIGHBORHOOD	164	17	9.5	2.8	14	PROTESTERS	127	1	127.0	443.9
15	DESPERATION	140	0	280.0	80.1	15	PEOPLE	121	50	2.4	8.5
16	EVENING	132	12	11.0	3.1	16	CHANGE	116	6	19.3	67.6
17	TOWN	128	23	5.5	1.6	17	REVOLUTION	106	110	1.0	3.4
18	SPOT	116	12	9.7	2.8	18	DEMONSTRATION	101	0	202.0	705.0
19	DOWN	111	0	222.0	63.5	19	DEMONSTRATIONS	98	0	196.0	685.0
20	REVOLUTION	110	106	1.0	0.3	20	RELATIONS	96	0	192.0	671.1

Рис. 5. Порівняння словосполучень 'quiet + іменник' та 'peaceful + іменник' у корпусі СОСА (сортування за частотою)

вання слів, словосполучень, граматичних конструкцій, багатозначності, омонімії, синонімії, антонімії, історичних змін в словниковому складі

мови. Перспективу подальших розвідок вбачаємо у дослідженні інших частин мови на основі даних цього корпусу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Демська-Кульчицька О. М. Деякі аспекти корпусної лінгвістики. Українська мова : наук.-теорет. журн. 2005. № 1. С. 44–51.
2. Демська-Кульчицька О. М. Основи національного корпусу української мови. Київ, 2005. 218 с.
3. Демська-Кульчицька О. М. Репрезентативність як ознака текстового корпусу. Українська мова. 2005. № 3. С. 100–107.
4. Жуковська В. В. Вступ до корпусної лінгвістики. Житомир: Вид-во ЖДУ ім. І. Франка, 2013. 142 с.
5. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. Донецьк, 2006. 188 с.
6. Шведова. М. О. Генеральний регіонально анотований корпус української мови (ГРАК) як інструмент дослідження лексико-граматичної варіативності. Людина. Комп'ютер. Комунікація : Збірник наукових праць, 2019. С. 145–148.
7. Широков В. А., Бугаков О. В., Грязнухіна Т. О. та ін. Корпусна лінгвістика. Київ: Довіра, 2005. 471 с.
8. Biber D. Using corpus-based methods to investigate grammar and use: some case studies on the use of verbs in English. 2001. P. 101–115.

9. Francis W. Language Corpora. Berlin –New York: Mouton de Gruyter, 1991. P.17–35.
10. Leech G. New resources, or just better old ones? *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 2007. P. 134–149.
11. Sinclair J. *Corpus, Concordance, Collocation* Oxford: Oxford University Press, 1991. 170 p.
12. Селіванова О. О. Корпусна лінгвістика. Сучасна лінгвістика: напрями та проблеми. Полтава, 2008. 712 с.
13. The Corpus of Contemporary American English. URL: <https://www.english-corpora.org/coca/>

UDC 81.33

DOI <https://doi.org/10.32782/tps2663-4880/2022.24.2.30>

THE DISCOURSE STUDIES OF COMPUTER-MEDIATED COMMUNICATION GENRES DEVELOPMENT IN MODERN LINGUISTICS

ДИСКУРСИВНІ ДОСЛІДЖЕННЯ РОЗВИТКУ ЖАНРІВ КОМП'ЮТЕРНО-ОПОСЕРЕДКОВАНОЇ КОМУНІКАЦІЇ В СУЧАСНІЙ ЛІНГВІСТИЦІ

Havrylenko K.M.,

orcid.org/0000-0001-9474-1990

Candidate of Pedagogical Studies,

Senior Teacher at the Department of English for Engineering № 2

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Prykhodko D.S.,

orcid.org/0000-0002-8276-5386

Teacher at the Department of English for Engineering № 2

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Modern linguistics works identify computer discourse analysis as one of the most vital scientific problems. The paper discusses various approaches to discourse studies and describes the linguistic and extralinguistic features of modern computer discourse. Communication in a computer-mediated discourse is an object of research and the linguistic features of scientific network discourse are described.

The paper describes the results of the concepts of computer-mediated communication studies and linguistic genres development based on the works of modern linguists. The results of the study show that digital technologies assist in the development of new types of communication, and change the old ones.

Discourse is a real speech element and a unit of verbal communication and could be regarded as a text, but the text is not a discourse but an abstract mental unit realized in discourse. Special attention is paid to the scientific and educational discourses, which are significantly influenced by computer communication discourse genre development and expand the scope of academic interaction.

Another aspect of the computer discourse study, described in the paper, is genre classifications research, which is analogical to the genres of traditional communication and includes general and new genres used in electronic media. The functional genre classification considers information, social, directive, presentational, aesthetic, and entertainment genres studied in the process of communicative interaction. The electronic genres' characteristics are dependent on communicative goals, mode of interaction, text forms, composition, and linguistic features.

Special attention was drawn to characteristics of the computer-mediated communication discourse research and further ways of the linguistic study of modern computer discourse were proposed.

Key words: computer discourse, computer-mediated communication, computer-mediated discourse, digital communication, digital discourse, sociolinguistics of computer communication, genres.

Сучасні лінгвістичні наукові праці визначають дискурсивний аналіз комп'ютерної лексики як одну з найактуальніших наукових проблем. У статті розглядаються різні підходи до вивчення дискурсу, описуються лінгвістичні особливості сучасного комп'ютерного дискурсу. Об'єктом дослідження визначається комунікація в комп'ютерно-опосередкованому дискурсі, описуються лінгвістичні особливості наукового мережевого дискурсу.

У статті досліджено різні концепції вивчення комп'ютерної комунікації та розвитку електронних жанрів на основі праць сучасних лінгвістів. Результати дослідження показують, що цифрові технології сприяють розвитку нових видів комунікації та змінюють старі.

Дискурс описується одночасно як мовленнєвий елемент і як одиниця вербальної комунікації, яку можна розглядати як текст, але текст позиціонується не як дискурс, а абстрактна ментальна одиниця, що реалізується