

КОРПУСНА ЛІНГВІСТИКА ТА ЇЇ ПЕРСПЕКТИВИ

Корольова А. В.,

*доктор філологічних наук, професор,
завідувач кафедри загального мовознавства та германістики,
Національний педагогічний університет
імені М. П. Драгоманова*



Шановна наукова громадо! Радо повідомляємо, що 5 грудня 2019 року на факультеті іноземної філології НПУ імені М. П. Драгоманова відбудеться майстер-клас у рамках проведення II Міжнародної науково-прикладної конференції (засобами відео-Інтернет-конференц-зв'язку) "Корпусна лінгвістика: діджиталізація в науці та освіті".

Перша науково-прикладна конференція, яка проводилася у 2018 році й згуртувала зацікавлених дослідників у цій сфері, продемонструвала, що корпусний підхід до вивчення мовних явищ має значні перспективи й переваги для якісного оброблення великих масивів мовних даних. Наразі триває робота в КНУ імені Тараса Шевченка щодо об'єднання наукових осередків – фахівців у царині прикладної лінгвістики – у вищих навчальних закладах та інших установах України в громадську організацію.

Запрошуємо всіх бажаючих взяти участь в авторському тренінгу Жуковської В. В. по роботі з корпусами, спрямованому на відпрацювання умінь із застосування емпіричних мовних даних у корпусно-керованих (corpus-driven) та корпусно-базованих (corpus-based) лінгвістичних дослідженнях.

NOTA BENE!

Для роботи з корпусами до початку конференції необхідно зареєструватися на платформах

English-Corpora.org та **Sketch Engine**.

Покрокова інструкція додається нижче.

Реєстрація на платформі English-Corpora.org

Крок 1. Перейти на <https://www.english-corpora.org/>

Крок 2.

The most widely used online corpora. [Overview](#), [search types](#), [looking at variation](#), [corpus-based](#)

The links below are for the online interface. But you can also download the corpora for use

Corpus	Download	# words	Di	Genre(s)
iWeb: The Intelligent Web-based Corpus		14 billion	6 co	Web
News on the Web (NOW)		8.6 billion+	20 cc	Web: News
Global Web-Based English (GloWbE)		1.9 billion	20 cc	Web (incl blogs)
Wikipedia Corpus		1.9 billion	(Various)	2014 Wikipedia
Corpus of Contemporary American English (COCA)		560 million	American	1990-2017 Balanced

Крок 3.

Please fill out the brief form below. Within one or two minutes, you will receive an email. Simply click on the link in that email, and you will be able to continue using the corpora. (Note: If you have already registered, please [log in](#) to see your profile)

Name (e.g. Mary | Smith)

Email address

Password

Country -- SELECT--

Category

- University professor: languages / linguistics 3 RESEARCHER
- Graduate student: languages or linguistics
- University professor: not languages / linguistics 2 SEMI-RESEARCHER
- Teacher: not university; not graduate student
- Graduate student: not languages or linguistics 1 NOT RESEARCHER
- Student (undergraduate)
- Other

I agree to the [Terms and Conditions](#) for this website

[PROBLEMS ??](#)

Крок 4.

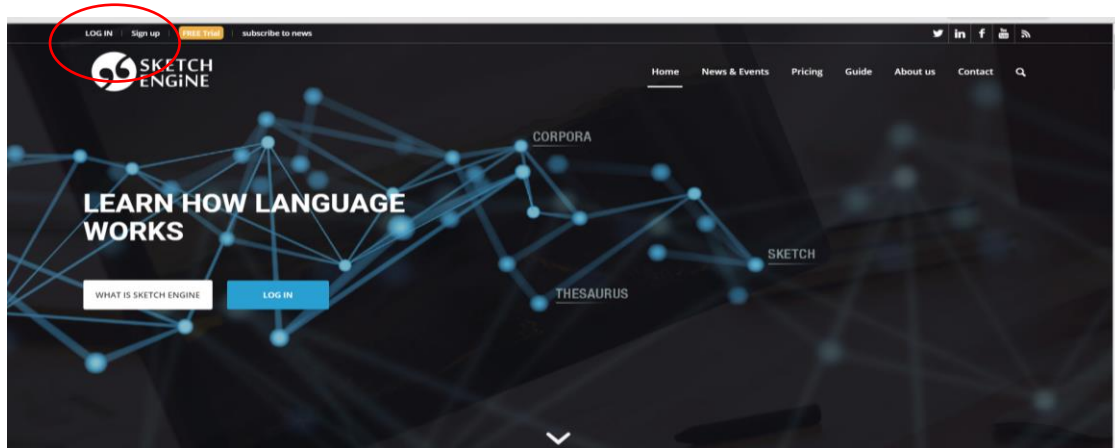
Крок 5. Після отримання листа, зайти у корпус і заповнити інформацію про користувача.

Крок 6. Зберегти / запам'ятати пароль для подальшої роботи з платформою.

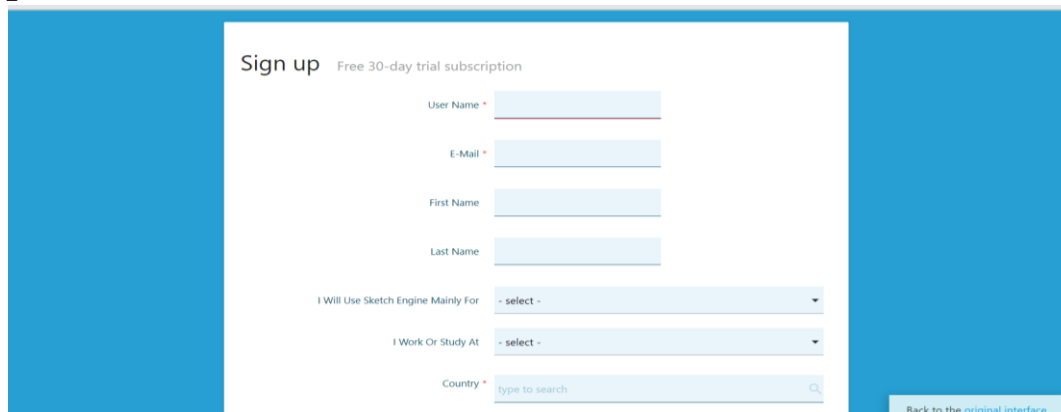
Реєстрація на платформі Sketch Engine (безкоштовна версія на 30 днів)

Крок 1. Зайти на <https://www.sketchengine.eu/>

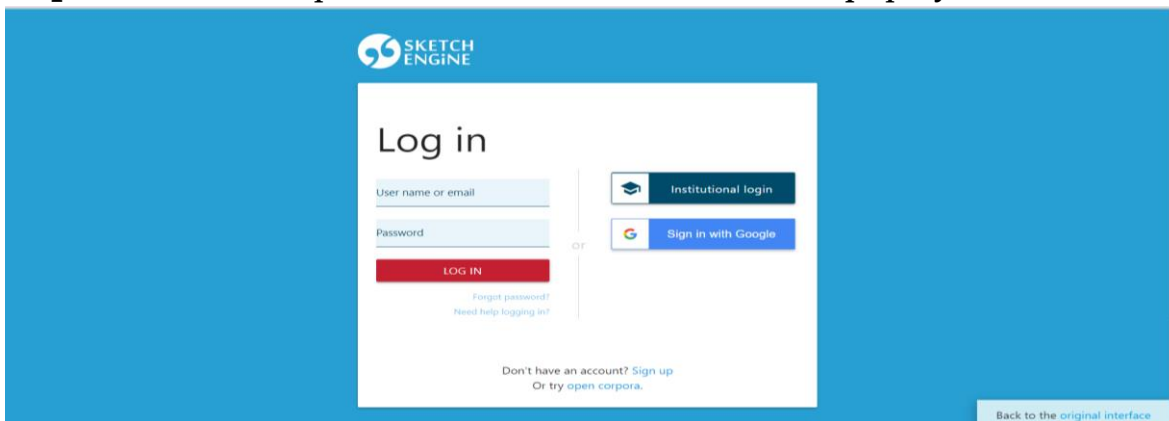
Крок 2.



Крок 3.



Крок 4. Після отримання листа, зайти на платформу



!!! Крок 5. Зберегти / запам'ятати пароль для подальшої роботи з платформою.

Бажаємо успіхів!

МАЙСТЕР-КЛАС



Жуковська Вікторія Вікторівна – кандидат філологічних наук, доцент, докторантка кафедри германської та фіно-угорської філології імені проф. Г. Г. Почепцова Київського національного лінгвістичного університету.

Т Р Е Н І Н Г

ПЕРША ЧАСТИНА

Лінгвістичні дослідження із залученням спеціалізованих пошукових систем корпусів текстів

Майстер-клас присвячено проведенню тренінгу по роботі з англійськомовними корпусами, зокрема випрацюванню умінь із застосування емпіричних мовних даних у корпусно-керованих (corpus-driven) та корпусно-базованих (corpus-based) лінгвістичних дослідженнях. Передбачено також роботу з корпусами для вивчення усного і писемного мовлення у синхронії та діахронії. Окрім того, заплановано огляд спеціалізованих корпусів англійськомовного академічного мовлення, а також технологійних можливостей застосування комп'ютерних програм для аналізу корпусних даних.

Майстер-клас орієнтований на молодих науковців, аспірантів, магістрантів, а також усіх, хто бажає опанувати інноваційний інструментарій та сучасні методи оброблення великих масивів мовних даних.

ДРУГА ЧАСТИНА

Практичне заняття зі студентами з дисципліни “Практичний курс англійської мови” із залученням корпусних можливостей

У перебігу заняття студенти й учасники майстер-класу ознайомляться з новим підходом до вивчення англійської мови як іноземної – навчанням на основі даних Data-driven learning. Головну увагу буде зосереджено на лінгводидактичних корпусних ресурсах, що можна застосовувати як для розроблення оригінальних навчальних матеріалів, так і для самостійного вивчення англійської мови.

На занятті буде продемонстровано можливості платформи BYU corpora, зокрема, розглянуто розміщені на платформі корпуси та корпусні ресурси. На практиці буде апробовано технологійне забезпечення корпусу сучасної американської англійської (Corpus of Contemporary American English) мови. Відбудеться навчання з коректності виконання запитів для пошуку в корпусі необхідної лінгвістичної інформації та подальшої інтерпретації отриманих даних. Відбудеться презентація інших лінгводидактичних корпусних ресурсів.

Заняття буде корисним для студентів і викладачів, які зацікавлені у корпусно-орієнтованому вивченні іноземної мови.

ТЕЗИ ДЛЯ ОБГОВОРЕННЯ ПРОБЛЕМНИХ ПИТАНЬ КОРПУСНИХ ТЕХНОЛОГІЙ

ПАРАЛЕЛЬНІ Й МУЛЬТИМОВНІ КОРПУСИ В ПЕРЕКЛАДАЦЬКІЙ ДІЯЛЬНОСТІ

Анохіна Т. О.,

*доктор філологічних наук, доцент,
доцент кафедри прикладної лінгвістики,
порівняльного мовознавства та перекладу,
Національний педагогічний університет
імені М. П. Драгоманова*

В технологічний час діджиталізація суттєво вплинула і на лінгвістику тексту, яка змушує змінити погляд на текст та його можливості. Сьогодні активно створюються електронні паралельні та мультимовні корпуси текстів. Важливо вміти застосовувати сучасні технології, щоб повною мірою використовувати лінгвістичний потенціал бітекстів та мультимовних термінологічних баз. За перспективним планом технологічного розвитку сучасні корпусні інструменти значно розширюють перекладацькі можливості.

Корпус – це контекстуальна гнучка платформа, частина електронної бази даних, яку має використовувати висококваліфікований сучасний перекладач. Корпус надає можливості для попереднього опрацювання текстів заданого спрямування (юридичні, медичні, технічні, художні тощо). Для удосконалення перекладу важливо знаходити тематичний перекладацький корпус, що може бути наявним як у вигляді монотекстів, так і бітекстів [4, с. 140].

Для ефективної роботи з корпусним матеріалом можна використовувати корпусні менеджери, а також інструменти для напіваавтоматичних систем перекладу [3, с. 341].

Існують різні напрями для роботи з корпусами, які задають використання певних інструментів. Деякі з них використовуються для роботи в оффлайн режимі: наприклад, робота в *AntConc* [2, с. 729], інші – мають динамічну інтернет-реалізацію: *COCA*, *Web corp*, а також корпус української мови.

ЛІТЕРАТУРА

1. Анохіна Т. О., Корольова А. В. Практичне застосування англо-українських бітекстів в перекладацьких студіях // *Філософія мови та нові*

тенденції в перекладознавстві й лінгвістиці: матеріали II Міжнародної науково-практичної конференції. Київ, 2019. С. 140–143.

2. Alghamdi A., Atwell E. Towards Comprehensive Computational Representations of Arabic Multiword Expressions // Europhras 2017 / Mitkov (Ed.), LNAI 10596. London: Springer International Publishing, 2017. P. 415–431.

3. Lawrence A. AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom // IEEE International Professional Communication Conference Proceedings. 2003. P. 729–737.

4. Veiga A. T. Using corpus linguistics tools to help translation students create technical glossaries // ICICTE 2016 Proceedings, 2016. P. 341–347.

ВИКОРИСТАННЯ БАГАТОМОВНИХ VS ОДНОМОВНИХ КОРПУСІВ У ТИПОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

Балабан О. О.,

кандидат філологічних наук, доцент,

доцент кафедри загального мовознавства і германістики,

Національний педагогічний університет

імені М. П. Драгоманова

У межах когнітивної лінгвістики як гетерогенної галузі дослідження мови, на думку В. Жуковської [1], визначаються окремі піднапрями, що різняться підходами до конструкції, експлуатації корпусів та аналізу корпусних даних. Одним із таких параметрів для аналізу корпусних даних є використання *багатомовних (multilingual) та одномовних (monolingual) корпусів*. Ще одним критерієм для розмежування типів корпусів є кількість мов, представлених у корпусі. Більшість корпусів репрезентують данні та варіативність на матеріалі однієї мови і є одномовними. Прикладом такого корпусу є The International Corpus of English, великий одномовний корпус, за допомогою якого можливо здійснювати контрастивні та типологічні дослідження національних варіантів англійської мови. Тимчасом багатомовні корпуси – це корпуси, що побудовані на матеріалі двох або більше мов. Вузьке розуміння багатомовного корпусу базується на тому, що він укладається принаймні трьома мовами, а корпуси, що містять дані двох мов, називаються двомовними (bilingual) корпусами. Багатомовні корпуси, у свою чергу, поділяються на порівняльні корпуси (comparable) та паралельні (parallel) корпуси. Порівняльний корпус містить

компоненти, що були вилучені за критерієм однакової вибірки текстів одного жанру та з однакових комунікативних сфер мов, що досліджуються в однаковий часовий період. Паралельний корпус має мову-джерело (L1) та її переклади (L2) (the CRATER, EMILLE, MULTEXT, P-ACTRES). Порівняльні та паралельні корпуси використовуються для проведення перекладацьких та контрастивних досліджень. Для порівняльних корпусів вагомим є суворе дотримання методології відбору текстів з досліджуваних мов: відібрані тексти повинні повністю збігатися за обсягом, жанром і часом створення. Для текстів-перекладів у паралельному корпусі ці параметри не є важливими, оскільки вони є точними перекладами текстів з мови-джерела. Цінною характеристикою паралельного корпусу є так зване вирівнювання (alignment) тексту-джерела та його перекладів, що передбачає анотування відповідностей між ними на рівні речення чи слова.

ЛІТЕРАТУРА

1. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ЖДУ імені І. Франка, 2013. 142 с.
2. Загнітко А., Данилюк І., Краснобаєва-Чорна Ж., Путіліна О., Ситар Г. Парадигмально-категорійні основи прикладної лінгвістики: монографія. Вінниця: "ТОВ Нілан-ЛТД", 2015. 472 с.

КОРПУС ATCOSIM ЯК ОСНОВА ВИЧЕННЯ ЛІНГВОКОМУНІКАТИВНИХ ОСОБЛИВОСТЕЙ РАДІООБМІНУ АНГЛІЙСЬКОМОВНОЇ ЦИВІЛЬНОЇ АВІАЦІЇ

Жуковська В. В.,

*кандидат філологічних наук, доцент,
докторантка кафедри германської і фіно-угорської філології
імені проф. Г. Г. Почепцова,*

Київський національний лінгвістичний університет

Бондар Х. В.,

магістрант,

Житомирський державний університет імені Івана Франка

Радіообмін цивільної авіації виявляється у двосторонньому обміні інформацією за допомогою різноманітних пристроїв радіозв'язку між екіпажем повітряного судна та диспетчерами або під час внутрішнього

радіообміну між членами одного екіпажу [1, с. 4–5]. Вивчення лінгвокомунікативних особливостей мовлення учасників радіообміну здійснюється на основі корпусу *ATCOSIM* (Corpus of Non-Prompted Clean Air Traffic Control Speech) [2], що містить транскрипти 50 годин записів реальних ситуацій англійськомовного радіообміну між диспетчерами та пілотами.

Результати аналізу 10078 затранскрибованих фрагментів виявили: 1) використання фонетичного алфавіту *ІКАО* для заміни літер звичайного алфавіту визначеними словами (*A – Alfa, D – delta*) та для передачі команд та дій за допомогою альфа-кодів (*Romeo Tango Bravo – повертаюсь на базу, Hotel Echo – виявлено вибухівку*); 2) вживання загальної лексики (*affirm – підтверджую*), спеціальної термінології сфери авіації (*roger – зрозумів*), іншомовної (*guten tag*) та розмовної лексики (*well i'm a little bit in stress*); 3) функціонування визначених модальних дієслів, зокрема, *can, could, should, need, have to* (*can you accept level three five zero*), часових форм (*Present/Future/Past Simple, Present Continuous* та *Present Perfect*) та частин мови (займенників, числівників, прикметників та прислівників); 4) вживання розповідних, питальних та спонукальних (наказ, порада, вимога, спонукання до дії) речень, запитань (загальних, спеціальних, альтернативних, розділових та запитання до підмета), відсутність альтернативних та розділових запитань; 5) використання елементів ввічливості (*sorry, please*), привітань (*hello, good afternoon*), а також виразів згоди та заперечення (*negative cleared level two nine zero*).

ЛІТЕРАТУРА

1. Олянюк П. В. Радионавигационные устройства и системы гражданской авиации. Москва: Транспорт, 1983. 585 с.

2. The *ATCOSIM* Corpus of Non-Prompted Clean Air Traffic Control Speech.
URL:

https://www2.spssc.tugraz.at/databases/ATCOSIM/HTMLdata/overview_dynamic.htm

КОРПУСНИЙ АНАЛІЗ У ТЕРМІНОЗНАВСТВІ

Вороніна-Пригодій Д. А.,

аспірантка кафедри загального мовознавства і германістики,

Національний педагогічний університет

імені М. П. Драгоманова

У сучасному мовознавстві корпусна лінгвістика є прикладною науковою дисципліною, що описує реальні мовні явища, їхні елементи та структуру на основі аналізу автентичних письмових або усних текстів, зібраних у лінгвістичні текстові корпуси. Корпус визначають як масив, зібрання правильно анотованих та структурованих письмових або усних висловлювань в електронному вигляді, що створюються для вирішення певних лінгвістичних завдань [2, с. 13-14].

Якщо розглядати корпусний аналіз у термінознавстві, то він спершу спрямований на пошук важливих термінів у тексті відповідної галузі, а потім, на пошук шаблонів знань (*knowledge patterns*), що свідчать про смислові зв'язки між цими термінами [3].

Для створення термінологічної бази знань (*Terminological Knowledge Base – ТКВ*) з метою подальшої термінологічної роботи необхідно сформувати корпус, який, як правило, визначається термінознавцем самостійно. Побудова корпусу є важливим етапом роботи, адже термінознавець повинен отримати спеціалізовані тексти з різних джерел. Ці тексти не повинні бути із будь-яким інформаційним наповненням, а саме з конкретної досліджуваної галузі [1, с. 2].

Існують критерії щодо вибору актуальних текстів:

- текст повинен бути якісним за змістом у спеціалізованій сфері знань;
- текст може бути обраним серед різних мов як один ключовий, оскільки одним із завдань термінології є визначення еквівалентів;
- тексти не повинні бути перекладами;
- тексти розподіляються за рівнем спеціалізації (складність тексту – написаний для експертів чи загальної аудиторії).

Також важливими показниками для якісного тексту є тип, стиль літератури, дані, що містяться в ньому, оцінка даних (автори чи репутація видавця) [1, с. 3].

Отже, тексти, обрані термінознавцями, міститимуть значну кількість шаблонів знань (*knowledge patterns*) відповідної галузі, оскільки вони заздалегідь будуть підготовлені до зіставлення смислових зв'язків з наявними в них термінами.

ЛІТЕРАТУРА

1. Agbago A., Barrière C. Corpus construction for terminology // Proceedings of the Corpus Linguistics Conference (July 14–17, 2005). Birmingham, United Kingdom, 2005. 14 p.
2. Lemnitzer L., Zinsmeister H. Korpuslinguistik: eine Einführung. Tübingen: Gunter Narr, 2006. 220 s.
3. Meyer I. Extracting Knowledge-rich Contexts for Terminography: A Conceptual and Methodological Framework // Recent Advances in Computational Terminology / eds D. Bourigault, C. Jacquemin, M. C. L'Homme. 2001. P. 279–302.

ОЦІНКА БЛИЗЬКОСТІ ДВОХ МОВ: РІШЕННЯ ПРО СПОРІДНЕНІСТЬ (за математичним алгоритмом Г. В. Раушенбаха)

Капранов Я. В.,

*кандидат філологічних наук, доцент,
докторант кафедри германської і фіно-угорської філології
імені проф. Г. Г. Почепцова,
Київський національний лінгвістичний університет*

На сучасному етапі розвитку історичної лінгвістики продовжують постулювати ідею застосування методів прикладної лінгвістики (як математичних, так і комп'ютерних (автоматизованих)) для виявлення мовних зв'язків без залучення дослідників (людських ресурсів) [див. пр.: 1; 7; 10]. Проте якщо прихильники підкреслюють дві основні переваги таких методів – *швидкість* та *повторюваність* [див. пр.: 4; 8], то скептики критикують *обґрунтованість* і *точність* цих методів [див. пр.: 6], а окремі вчені ставляться до них з обережністю [див. пр.: 3; 5; 9]. Основна причина, що ускладнює оцінку результатів, – міркування про фонетичні (звукові) відповідники та приховані рішення про спорідненість. Така непрозорість дещо ускладнює вдосконалення алгоритмів.

Розглянемо математичний алгоритм Г. В. Раушенбаха [2], заснований на розрахунку близькості двох мов і названий дослідником бінарним порівнянням, коли ступінь близькості об'єктів $X = \{x_1, x_2 \dots x_n\}$ і $Y = \{y_1, y_2 \dots y_n\}$, що описані дихотомічними ознаками x_i і y_i , прийнято виражати у вигляді функції чотирьох змінних:

$$S = F(a, b, c, d),$$

де a – число ознак, що не фіксуються у X і Y одночасно, d – число ознак, що збігаються, b (або c) – число ознак, що фіксуються в X , але не фіксуються в Y (або навпаки). Сума $(a + b + c + d)$ дорівнює загальному числу ознак n .

Дослідник пропонує такі формули для оцінки близькості двох мов:

$$S_1 = \frac{a+d}{a+b+c+d} ;$$

$$S_2 = \frac{2(a+d)}{2(a+d)+b+c} ;$$

$$S_3 = \frac{a+d}{a+d+2(b+c)} ;$$

$$S_4 = \frac{d}{a+b+c+d} ;$$

$$S_5 = \frac{d}{b+c+d} ;$$

$$S_6 = \frac{a}{a+b+c} ;$$

[2]

Ступінь близькості $S_1 - S_6$ може бути представлений у вигляді функцій від $v(X)$, $v(Y)$, $v(X \cap Y)$ або $v(X \cup Y)$. Це можна пояснити тим, що близькість об'єктів X і Y визначено наявністю в них загальних ознак, інформаційна цінність яких ($v(X \cap Y)$) повинна бути співвіднесена з інформацією про об'єкти, що розглядаються окремо ($v(X)$, $v(Y)$) або в поєднанні ($v(X \cup Y)$). Подібність об'єктів може бути виміряна не лише за наявністю, але й за відсутністю в них одних і тих же ознак – “подібність за антипатією” [1, с. 56].

Поєднавши і осмисливши зазначене вище, що було представлено у формулах $S_1 - S_6$, В. М. Поляков пропонує загальну формулу для оцінки близькості двох мов:

$$S_7 = \frac{R1*a + R4*d}{R1*a + R2*b + R3*c + R4*d} ,$$

де $R1 \dots R4$ – коефіцієнти, що визначають внесок у загальний ступінь

одночасної наявності ознак (параметр a), одночасну відсутність ознак (d) або протиріч (b і c) в об'єктах, що порівнюються [1, с. 56].

Рішення про спорідненість мов (як двох, так і більше) потребують їх глибокої аргументації не лише в плані залучення широкої експланаторної бази різних наук, а власне прикладних механізмів для оцінки близькості / віддаленості, а відтак, і зорієнтувати сучасну історичну лінгвістику до залучення здобутків точних наук.

ЛІТЕРАТУРА

1. Поляков В. Н., Соловьев В. Д. Компьютерные модели и методы в типологии и компаративистике: монография. Казань: Казанский государственный университет им. В. И. Ульянова-Ленина, 2006. 208 с.
2. Раушенбах Г. В. Меры близости и сходства // Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1985. С. 169–203.
3. Campbell L. Comment on: Automated dating of the world's language families based on lexical similarity // *Current Anthropology*. 2011. 52. P. 866–867.
4. Embleton S. Lexicostatistics / glottochronology. From Swadesh to Sankoff to Starostin to future horizons // *Time depth in historical linguistics* / eds Renfrew C., McMahon A., Trask L. Cambridge: The McDonald Institute for Archaeological Research, 2000. P. 143–165.
5. Greenhill S. J. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics*. 2011. 37(4). P. 689–698. Doi: 10.1162/COLI_a_00073 16.
6. Holm H. J. The new arboretum of Indo-European "trees" // *Journal of Quantitative Linguistics*. 2007. 14 (2–3). P. 167–214. Doi: 10.1080/09296170701378916
7. Holman E. W., Wichmann S., Brown Ch., Velupillai V., Muller A., Bakker D. Explorations in automated lexicostatistics. *Folia Linguistica*. 2008. 20 (3). P. 116–121.
8. McMahon A., McMahon R. *Language classification by numbers*. Oxford: Oxford University Press, 2005.
9. Sidwell P. Comment on: Automated Dating of the World's Language Families Based on Lexical Similarity // *Current Anthropology*. 2011. 52. P. 869–870.
10. Wheeler W. C., Whiteley P. M. Historical linguistics as a sequence optimization problem: the evolution and biogeography of Uto-Aztecan languages. *Cladistics*. 2015. 31. P. 113–125. Doi: 10.1111/cla.12078

ЗАСТОСУВАННЯ ЗАКОНУ ЦИПФА-МАНДЕЛЬБРОТА ДЛЯ РАНГОВО-ЧАСТОТНОГО РОЗПОДІЛУ ДАВНЬОАНГЛІЙСЬКИХ ЗАПОЗИЧЕНЬ (на матеріалі героїчної поеми “Беовульф”)

Карлова В. О.,

*аспірантка кафедри загального мовознавства і германістики,
Національний педагогічний університет
імені М. П. Драгоманова*

Сучасна лінгвістика спрямовує свої дослідження на отримання об'єктивних і точних результатів [2], які потребують застосування традиційних методів мовознавства поряд із математичними [1]. Серед таких методів варто назвати математичний закон Ципфа, який був доповнений на початку 50-х рр. ХХ ст. Б. Мандельбротом [4].

Звертаючись до закону Ципфа-Мандельброта, тобто до “закону частоти слів”, який є емпіричною закономірністю розподілу частоти слів природної мови, слід навести його формулу обчислення: $f * r = c$, де f – частота слова у тексті; r – ранг (порядковий номер слова) у списку; c – емпірична постійна величина [3]. Даний закон ефективно було використано під час дослідження запозичень у давньоанглійській поемі “Беовульф” для вимірювання частоти їх вживання та рангу в тексті епосу. Це дало змогу зробити попередні припущення з приводу лексичних лакун у мові давніх англів, для заповнення яких використовували запозичення.

ЛІТЕРАТУРА

1. Гребенщикова А. В. Основы квантитативной лингвистики и новых информационных технологий: учебное пособие. 2-е изд. М.: Флинта, 2015. 152 с.
2. Рябцева Н. К. Язык и естественный интеллект: монография. М.: Академия, 2005. 639 с.
3. Zipf G. K. Human behavior and the principle of least effort. Addison-Wesley Press. 1949. 573 p.
4. Zipf G. K. The Psychobiology of Language. Houghton-Mifflin, 1935. 336 p.

ВИКОРИСТАННЯ ІНСТРУМЕНТІВ КОРПУСНОЇ ЛІНГВІСТИКИ ДЛЯ ФОРМУВАННЯ ЛЕКСИЧНОЇ КОМПЕТЕНЦІЇ

Качанов В. І.,

кандидат військових наук,

доцент кафедри загального мовознавства і германістики,

Національний педагогічний університет

імені М. П. Драгоманова

У сучасному процесі вивчення іноземних мов активно використовуються інноваційні методи і підходи, одним з яких є корпусний [1; 2]. Корпусний підхід у розвитку лексичної компетенції спирається на використання найбільш вживаних слів та комбінацій слів у словосполученнях і реченнях [3]. Запас сталих виразів дозволяє покращити як рівень розуміння іноземної мови, так і сприяє збільшенню швидкості розмовного мовлення [3]. Знання сталих виразів допомагає визначити, яка лексична одиниця переважно використовується у тому чи іншому словосполученні та як вони об'єднуються в реченнях. При цьому, навіть студенти з високим рівнем володіння іноземною мовою мають певні труднощі у використанні сталих виразів [3]. Одним із рішень є "...автоматизація словосполучень, що досягається усвідомленням студентами детальної інформації щодо зразків їх використання, способів об'єднання в реченнях, рівня формальності та частотності вживання" [3]. В цьому відношенні корпусний метод є дуже важливим, адже допомагає дослідити словосполучення та сталі вирази не тільки кількісно, а й надати інформацію щодо вживання ключових слів та частотності їх використання, а також забезпечити студентів великим обсягом автентичної інформації у структурованому вигляді [3].

ЛІТЕРАТУРА

1. Bennett G. Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers. University of Michigan Press ELT, 2010. P. 9–11.
2. Reppen R. Using corpora in the language classroom. Cambridge: Cambridge University Press, 2010. P. 14–15.
3. Tony McEnery and Richard Xiao. What Corpora Can Offer in Language Teaching and Learning. URL: [https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/\[...\]](https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/[...])

СПЕЦІАЛІЗОВАНИЙ КОРПУС АКАДЕМІЧНОГО МОВЛЕННЯ ЯК ОСНОВА СОЦІОЛІНГВІСТИЧНОГО ДОСЛІДЖЕННЯ

Комаренко В. В.,

Житомирський державний університет імені Івана Франка

З-поміж наявних корпусів академічного мовлення (Британський академічний корпус (*The British Academic Spoken English Corpus: BASE*), Академічний текстовий корпус (*The Reading Academic Text Corpus*), Корпус усного професійного англійського мовлення (*The Corpus of Spoken Professional American-English*), Корпус професійного англійського мовлення (*The Corpus of Professional: PERC English*)), Мічиганський корпус усного англійського академічного мовлення (*The Michigan Corpus of Academic Spoken English (MICASE)*) [1] є одним із найкращих.

Корпус створений дослідниками та студентами Мічиганського інституту англійської мови. MICASE містить затранскрибовані матеріали 200 годин записів академічних подій з 4 предметних галузей: гуманітарні (*Humanities and Arts*), соціальні (*Social Sciences*), біологічні (*Biological and Health Sciences*), фізичні науки (*Physical Sciences*). Загальний обсяг корпусу становить 1,848,364 слів, що включають 152 фрагменти з широкого спектру академічних жанрів: лекцій, семінарів, дискусій, лабораторних робіт, семінарів, колоквиумів, інтерв'ю, студентських презентацій, захистів дисертацій загальним обсягом, що відбувалися в Мічиганському університеті у період з 1997 по 2002 рік. Представлені у корпусі мовленнєві події характеризуються різною протяжністю у часі від 19 до 178 хвилин та різним обсягом слововживань (від 2,805 до 30,328).

Корпус оснащений зручним інтерфейсом, що дозволяє здійснювати автоматизований пошук за різними соціолінгвістичними параметрами: гендерна приналежність мовців (мовлення чоловіків та жінок); тип події (консультації, колоквиуми, захист дисертацій, заняття-опитування, інтерв'ю, лабораторні заняття, довготривалі та короткотривалі лекції, зустріч з керівництвом факультету або студентською радою, офісні години, семінари, семінари дослідницьких груп, студентські презентації, службові зустрічі, тури-екскурсії бібліотекою, музеєм або студентським містечком); предметна сфера (гуманітарні, соціальні, біологічні, фізичні науки); ступінь монологічності та інтерактивності (високо інтерактивні, переважно інтерактивні, високо монологічні, переважно монологічні та змішані), академічна роль мовців (мовлення студентів молодших, старших курсів, випускників, викладачів та співробітників). Така багатоаспектна розмітка корпусу MICASE надає можливості для проведення ґрунтовного соціолінгвістичного дослідження.

ДЖЕРЕЛА ІЛЮСТРАТИВНОГО МАТЕРІАЛУ

1. Michigan Corpus of Academic Spoken English (MICASE). Dr. Ute Römer, 2010. URL: <http://www.helsinki.fi/varieng/CoRD/corpora/MICASE/index.html>

КОРПУСНО-БАЗОВАНІЙ АНАЛІЗ ФРАЗЕОЛОГІЗМІВ НА ПОЗНАЧЕННЯ ЕМОЦІЙ В АНГЛІЙСЬКІЙ ТА УКРАЇНСЬКІЙ МОВАХ

Жуковська В. В.,

кандидат філологічних наук, доцент,
докторантка кафедри германської і фіно-угорської філології
імені проф. Г. Г. Почепцова,
Київський національний лінгвістичний університет

Кравчук М. В.,

магістрант,
Житомирський державний університет імені Івана Франка

Дослідження фразеологічних одиниць (далі – ФО) на позначення емоцій здійснювався на основі текстових корпусів сучасних англійської (СОСА) [4] та української (ГРАК) [2] мов.

Корпусний аналіз наповненості тематичних груп ФО на позначення позитивних (*a happy go lucky, pride and joy; щасливий час, душа радіє*), негативних (*worried to death, rage out of control; лити сльози, серце завмирає від страху*) та нейтральних (*time works wonders; з доброго дива*) емоцій, виокремлених зі словників “Oxford Dictionary of English Idioms” [3] та “Словника фразеологізмів української мови” [1], дозволив виявити такі особливості функціонування досліджуваних одиниць: 1) не всі ФО аналізованої семантики активно вживаються у мовленні, про що свідчить менша кількість зафіксованих ФО у корпусі відповідної мови порівняно з кількістю ФО, відібраних з лексикографічних джерел; 2) в обох мовах найбільш наповненою є тематична група ФО на позначення негативних емоцій, а найменш частотними є ФО на позначення нейтральних емоцій; 3) досліджувані ФО найбільше вживаються в писемному мовленні у художніх та публіцистичних текстах.

ЛІТЕРАТУРА

1. Білоноженко В. М. Словник фразеологізмів української мови. Київ: Наук. думка, 2003. 786 с.
2. Генеральний регіонально анотований корпус української мови. URL: http://www.parasolcorpus.org/bonito/run.cgi/first_form
3. Ayto J. Oxford Dictionary of English Idioms. Oxford: Oxford University Press, 2009. 512 p.
4. Corpus of Contemporary American English. URL: <https://www.english-corpora.org/time>

ПЕРСПЕКТИВИ ПАРАЛЕЛЬНИХ КОРПУСІВ В УКРАЇНІ

Леміш Н. Є.,

*доктор філологічних наук, доцент,
завідувач кафедри прикладної лінгвістики,
порівняльного мовознавства та перекладу,
Національний педагогічний університет
імені М. П. Драгоманова*

Глобалізація світу та інтеграція України у світовий простір детермінують зростання потреби українського суспільства у професійних перекладачах. Реагуючи на цей виклик, ціла низка закладів вищої освіти, серед яких Київський національний лінгвістичний університет, Київський національний університет імені Тараса Шевченка, Національний педагогічний університет імені М. П. Драгоманова та ін., переглядає методи навчання майбутніх перекладачів, фокусуючи увагу на розробці та застосуванні інформаційних технологій своїми вихованцями. *Найактуальнішим* у цьому зв'язку постає завдання створення паралельних корпусів для користування викладачами та студентами спеціальності “Переклад”.

Виконання вищезазначеного завдання є можливим з огляду на досвід наших іноземних колег, які вже досягли певних успіхів на цьому шляху. По-перше, це перехід від корпусного підходу типу *corpus-assisted* (пошук прикладів для ілюстрування дефініцій) до *corpus-driven* (пояснення різних дефініцій на основі ілюстрацій з корпусів задля обґрунтування попередньо сформульованих теорій) [1, с. 2214; 2, с. 290, 292–293]. По-друге, на зміну статичних колекцій текстів приходять інструменти для створення динамічних корпусів, що відкриває перспективи для вирішення індивідуальних / спеціальних запитів [3, с. 153]. По-

третє, кількість паралельних корпусів різних мов у світі дуже швидко зростає (наприклад: MBS – бельгійська, французька, нідерландська мови; OpenSubtitles2011 – більше 50 мов; TEP – перська та англійська мови; WikiSource – шведська та англійська мови тощо), серед яких OPUS є найбільшим, включаючи більше 3800 мовних пар [1, с. 2214–2215].

Таким чином, років через п'ять в інформаційному просторі українських закладів вищої освіти очікується поява принаймні кількох паралельних корпусів, які полегшать не лише пошук окремих лексичних / фразеологічних одиниць або синтаксичних структур у певних контекстах, але й дадуть змогу визначати найефективніші типи перекладацьких трансформацій для відтворення у цільовій мові конкретних мовних феноменів мови-оригіналу.

ЛІТЕРАТУРА

1. Tiedemann J. Parallel Data, Tools and Interfaces in OPUS // Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), May 2012, Istanbul, Turkey, 2012. P. 2214–2218. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf. (in English).
2. Tissari H., Vanhatalo U., Siirainen M. From Corpus-Assisted to Corpus-Driven MSM Explications: the Case of Finnish *VIHA* (Anger, Hate) // *Lege artis. Language yesterday, today, tomorrow*, 2019. Vol. IV. No. 1. P. 290–334.
3. Vintar Š. Corpora in Translator Training and Practice: A Slovene Perspective // *Incorporating Corpora. The Linguist and the Translator*. Clevedon: Multilingual Matters Ltd., 2008. P. 153–167.

ФОРМУВАННЯ МАСИВІВ ПЕРЕКЛАДНИХ ЕКВІВАЛЕНТІВ: КОРПУСНИЙ ПІДХІД

Матвеева С. А.,

*кандидат філологічних наук, професор,
професор кафедри прикладної лінгвістики,
порівняльного мовознавства та перекладу,
Національний педагогічний університет
імені М. П. Драгоманова*

Завданням будь-якого двомовного словника є, перш за все, надання якомога більшої кількості варіантів перекладу, тобто презентація максимально повного масиву перекладацьких еквівалентів кожної одиниці. Відсутність

можливості врахування контексту й дистрибуції мовної одиниці перетворює словник на суто довідкове джерело, дані якого можливо застосовувати не для перекладу, а тільки для уточнення змісту одиниці, що перекладається. Корпус паралельних текстів у певному розумінні “пропонує сукупність перекладацьких стратегій, до яких вдавалися перекладачі, які раніше стикалися з проблемами, подібними до тих, що викликали необхідність звернення до паралельного корпусу” [1, с. 11]. Використання корпусу паралельних текстів дозволяє не тільки формувати актуальні масиви перекладних еквівалентів, але й виявляти закономірності в перекладі певної мовної одиниці засобами іншої мови з подальшим застосуванням таких закономірностей до нових одиниць перекладу, уточнення й збагачення реєстру перекладацьких еквівалентів у двомовних словниках.

ЛІТЕРАТУРА

1. Zanettin F. Corpora in Translation Practice // Language Resources for Translation Work and Research. 2002. P. 10-14.

ВИКОРИСТАННЯ ПАРАЛЕЛЬНИХ КОРПУСІВ У ВИВЧЕННІ ІНОЗЕМНИХ МОВ

Ружицька А. О.,

*аспірантка кафедри загального мовознавства і германістики,
Національний педагогічний університет
імені М. П. Драгоманова*

З розвитком новітніх технологій з'являються нові способи обробки мовного матеріалу. Це дало поштовх до зародження такого розділу мовознавства, як корпусна лінгвістика, одним з напрямків якої є створення та використання паралельних корпусів. Окрім оригінальних текстів, вони містять їх переклади іншими мовами та сегментовані за реченнями або абзацами. Застосування паралельних корпусів має велике значення для перекладацької діяльності та для вивчення іноземних мов, тому що вони можуть слугувати наочним прикладом для навчання методам та прийомам перекладу. Корпус паралельних текстів дає змогу проаналізувати мовну одиницю та варіанти її перекладу із залученням великого масиву текстової та екстралінгвістичної інформації з одночасним зануренням у контекст [2, с. 185]. Це пов'язане з тим, що контекстуальний переклад лексичної одиниці може суттєво відрізнитися від словникового. Крім

того, приклади, які взяті з корпусів, є більш ефективними, адже дозволяють аналізувати реальні слововживання в природному мовному середовищі та надають більш точний та деталізований опис мовних одиниць, що сприяє опануванню мови на належному рівні [1, с. 278].

Отже, застосування паралельних корпусів у вивченні іноземних мов сприяє розширенню словникового запасу, формуванню навичок лінгвістичного аналізу та підвищує рівень перекладацьких компетентностей студентів.

ЛІТЕРАТУРА

1. Маслова Т. Б. Інструменти корпусної лінгвістики у навчанні і дослідженні іноземних мов. Smart-освіта: ресурси та перспективи: тези доп. II Міжнар. наук.-метод. конф. (м. Київ, 23 лист. 2016 р.). Київ, 2016. С. 277–280.

2. Матвеева С. А. Корпус паралельних текстів як інструмент перекладача. Система і структура східнослов'янських мов. 2016. № 10. С. 183–189.

ДОСЛІДЖЕННЯ СЕМАНТИЧНИХ ПРОЦЕСІВ У ЛЕКСИЦІ В КОРПУСАХ ТЕКСТІВ

Юрченко Н. С.,

аспірантка кафедри загального мовознавства і германістики,

Національний педагогічний університет

імені М. П. Драгоманова

Звернення до корпусних текстів дозволяє проаналізувати питання лексичної семантики за допомогою засобів програмного забезпечення. Це робить процес збирання, опрацювання, збереження та аналізу даних швидшим та простішим. Під час вивчення мовних явищ у синхронії та діахронії застосування корпусного методу буде корисним для аналізу лексичних одиниць, варіантів їх функціонального використання, у визначенні частотності вживання слів. Характерними особливостями застосування даного методу можна визначити системність охоплення більшої кількості одиниць, аналіз динамічних змін у мові, закономірності її розвитку.

У дослідженнях багатозначних слів завдяки корпусним даним можна проаналізувати фактичне вживання лексичних одиниць в різних комунікативних ситуаціях, визначити їх найхарактерніші значення та особливості розуміння різними мовцями. Це дозволяє встановити набір семантичних компонентів, що поєднані відношеннями різних типів. Значення слів при цьому аналізується завдяки контекстним значенням, у конкордансах. Для визначення семантики

багатозначних слів аналізується контекстне вживання, валентність слів та частотність їх вживання.

Розвиток корпусних досліджень є важливим для створення автоматизованих лексикографічних систем та потребує подальших розробок та вдосконалення.

ЗМІСТ

Корольова А. В.	Корпусна лінгвістика та її майбутнє.	4
-----------------	--	---

МАЙСТЕР-КЛАС

Жуковська В. В.		
ПЕРША ЧАСТИНА	Лінгвістичні дослідження із залученням спеціалізованих пошукових систем корпусів текстів	6
ДРУГА ЧАСТИНА	Практичне заняття зі студентами з дисципліни “Практичний курс англійської мови” із залученням корпусних можливостей	7

Тези для обговорення проблемних питань корпусних технологій

Анохіна Т. О.	Паралельні й мультимовні корпуси в перекладацькій діяльності	8
Балабан О. О.	Використання багатомовних vs одномовних корпусів у типологічних дослідженнях	9
Жуковська В. В., Бондар Х. В.	Корпус <i>Atcosim</i> як основа вивчення лінгвокомунікативних особливостей радіообміну англійськомовної цивільної авіації	10
Вороніна-Пригодій Д. А.	Корпусний аналіз у термінознавстві	12
Капранов Я. В.	Оцінка близькості двох мов: рішення про спорідненість (за математичним алгоритмом Г. В. Раушенбаха)	13
Карлова В. О.	Закон Ципфа-Мандельброта для рангово-частотного розподілу давньоанглійських запозичень (на матеріалі героїчної поеми “Беовульф”)	16
Качанов В. І.	Використання інструментів корпусної лінгвістики для формування лексичної компетенції	17
Комаренко В. В.	Спеціалізований корпус академічного мовлення як основа соціолінгвістичного дослідження	18
Жуковська В. В., Кравчук М. В.	Корпуснобазований аналіз фразеологізмів на позначення емоцій в англійській та українській мовах	19
Леміш Н. Є.	Перспективи паралельних корпусів в Україні	20

Матвеева С. А.	Формування масивів перекладних еквівалентів: корпусний підхід	21
Ружицька А. О.	Використання паралельних корпусів у вивченні іноземних мов	22
Юрченко Н. С.	Дослідження семантичних процесів в лексиці в корпусних текстах	23