

УДК 811.111Т373.2

**В. Ю. Неклесова**

кандидат філологічних наук, доцент кафедри іноземних мов  
гуманітарних факультетів Одеського національного  
університету ім. І.І. Мечникова e-mail:  
[neklesova.valerie@gmail.com](mailto:neklesova.valerie@gmail.com)

## ФОРМУВАННЯ УКРАЇНСЬКОГО ОНОМАСТИЧНОГО КОРПУСУ

*У статті розглядається питання утворення українського ономастичного корпусу. Розглянуто вже існуючі корпуси та ті корпуси, де присутні дані з української мови.*

**Ключові слова:** ономастика, український ономастичний корпус.

Мета статті полягає в огляді існуючого стану корпусів, які містять дані з української мови на теперішній час. Завдання нашої статті вбачається у розгляді можливостей та перспектив утворення такого корпусу текстів для розв'язання завдань, які ставить перед собою ономастика. Об'єктом є можливість фузії корпусної лінгвістики та української ономастики. Предметом є засоби формування та аотації Українського ономастичного корпусу.

Корпусна лінгвістика в Україні має не дуже велику історію і є майже на початку свого розвитку. Так, в огляді корпусних систем слов'янських мов, науковець Т. І. Резникова зазначає, що для української мови ще не створено загальнодоступного корпусу [7, с. 442]. Це свідчить не стільки про відсутність взагалі корпусних досліджень та праць в україністиці, скільки про те, що корпусній лінгвоукраїністиці потрібна популяризація та розширення ресурсів. **Національний проект корпусу української мови** [16] відкрито для багатьох дослідників уже 10 років. Цей корпус є анованим граматично, лексично та гендерно (можна обрати автора за статтю). Існує багатомовний проект **The Tatoeba Project** [17], де також представлено українську мову, яка знаходиться на 17 місці за статистикою сайту (23290 речень), що є непоганим результатом. Це усі дані про стан корпусної лінгвоукраїністики, про які нам відомо. Отже, нагальна потреба корпусної україністики є в розширенні корпусів української мови.

Сьогодні не існує корпусу, який би базувався тільки на власних назвах української мови. Такий стан речей не може бути задовільним, адже ономастичні студії в Україні мають велику історію, прикро, що багато з цих досліджень на теперішній час недоступні онлайн ані для дослідників, ані для широкого кола читачів, які захоплюються історією та походженням власних назв. На думку Р. Мейера корпус — це зібрання текстів або частин текстів, які можна проаналізувати за допомогою певних лінгвістичних методик [10, с. xi].

Існують топонімічні комп'ютерні банки даних: топоніміки півночі Росії TORIS [18] та Ленінградської та Псковської Областей KNAB (на жаль, зараз немає доступу до цієї бази даних онлайн). Які стали базою для досліджень даних Інгерманландії О. В. Дмитрієва [2]. Мова йде лише про топонімію, інші розряди власних назв є практично неохопленими Інтернетом.

Стандартне визначення лінгвістичного корпусу — це колекція текстів, яка зберігається в електронних базах даних. [9, с. 48-49]. Дещо розширеним визначенням послуговується О. М. Демська: перетворена на електронну форму, стандартно організована, програмно опрацьована, репрезентативна вибірка текстів певної мови (мов), призначена для наукового та практичного її (їхнього) вивчення [1, с. 9]. Більш широку дефініцію пропонує В. В. Риков: «корпус текстів — це певним чином організована множинність, елементами якої є тексти» [8]. Як бачимо, наведені дефініції збігаються в тому, що корпус — це зібрання текстів, які було складено з метою лінгвістичного осмислення. Для нашого проекту корпусу, основою є зібрання текстів, які містять власні назви. Такий корпус буде спеціалізованим, тобто таким, який працює лише з певним лінгвістичним явищем або створений для якогось дослідницького проекту, наприклад спеціалізованим вважається корпус на базі видання TIME [19], що містить тільки зразки, вилучені з фонду цього видання.

У мережі Інтернет є доволі розгалужені корпуси на багатьох мовах, досвід їх складання, конструювання та складу може допомогти для організації Українського ономастичного корпусу (далі Корпус або Ономастичний корпус). Еталонами корпусів стали такі з англійської мови як **The Brown University Standard Corpus of Present-Day American English** (or just **Brown Corpus**) [14], **The British National Corpus (BNC)** [15], **The American National Corpus (ANC)** [12]. Вони вважаються одні-

ми з найчисленніших колекцій текстів англійської мови, де електронний банк постійно поповнюється, а розмітка є найпрактичнішою.

Проект Ономастичного корпусу полегшить багато завдань. Можна буде провести не тільки якісний, але й спиратися на кількісний аналіз онімів не тільки у певних текстах певних авторів, а відстежити їх по жанрах та по засобах узусу. Як зазначає В. О. Плунгян корпусна лінгвістика є взагалі «ноюю ідеологією, яка орієнтує дослідника на *текст* як головний об'єкт теоретичної рефлексії» [5, с. 14].

Використовуючи Ономастичний корпус дослідник зможе оминати певну суб'єктивність. Зараз дослідники спираються у більшості на індивідуально відібрані тексти або приклади, що може привносити певну упередженість у процес дослідження. Таку думку висловлює Н. В. Перцов: «мовну інтроспекцію дослідника як носія мови потрібно верифікувати завдяки зверненням до корпусів та опитуванню інформантів» [4, с. 327]. Ономастичний корпус допоможе оминати такий стан речей. Ономастичні дослідження виходять із емпіричного матеріалу, персональної картотеки дослідника, якщо ми змогли б об'єднати їх, така персональна картотека не була би, так би мовити, «одноразовою», навпаки її можна було би використати ще раз та зробити доступною для інших студій.

Корпус набагато облегшить працю над словниками певних письменників або навіть жанрів, допоможе в студіях зв'язків власних назв із іншими лексемами. З даних Ономастичного корпусу можна буде знайти інформацію про конотонімізацію та простежити апелятивізацію власних назв. За допомогою Корпусу можна буде простежити динаміку вживання онімів, також виявити ступінь вживаності прецедентних власних назв та міру наскільки певні прецеденти оніми є поширеними у різних верств населення (вік, соціальний статус, регіон). Ще одна можливість: виявлення та фіксація онімів-неологізмів, та виявлення чи втрата нових значень та конотацій у вже існуючих.

Доречно було б організувати пошук по Ономастичному корпусу за декількома ключами, декількома засобами: ключове слово, ономастичний розряд, якщо наявна міжонімна омонімія, за лемою або автором тексту, якщо ми говоримо про мову певного митця, стилем тощо.

Анотацію корпусу можна проводити на багатьох рівнях та багатьма засобами: на фонологічному, морфологічному, лексичному та синтаксичному рівнях. На рівні дискурсу корпус можна анутовати для

того, щоб показати анафоричні зв'язки [11, с. 158]. Також можливо зосередитися на прагматичному або стилістичному аспектах.

Архітектура Ономастичного корпусу складається з декількох шарів: художній, публіцистика, Інтернет-спілкування, розмовних, фольклорний, діалектний (зафіксувати різноманітність української мови в її регіональних варіантах). Головні дві частини — це писемні тексти та усні тексти. Писемні тексти можуть містити статті з газет та журналів, цитати з художніх творів. Усний варіант корпусу: реклама, розмови, колекція фільмів або телесеріалів, випуски новин телеканалів, тексти сучасних та народних пісень, інтерв'ю зірок, парламентських діячів та громадян. Електронна комунікація включає цитати з форумів, блогів. Останньому шару буде приділено особливу увагу, адже це новий тип спілкування. Електронна комунікація є особливим явищем, вона є сплавом писемного та усного в мові, на думку Е. В. Рахліної — це «дуже важливий шар мови, саме тут відбуваються інноваційні процеси, дещо інші, ніж в усному мовленні» [6, с. 13]. Хронологічні границі корпусу власних назв планується зробити широкими: від перших пам'яток до теперішнього часу.

Дозволимо собі скласти приблизний план, за яким планується скласти корпус: по-перше, ми не будемо виокремлювати типи власних назв, адже ще й досі не обговорено багато аспектів класифікації власних імен. Навіть серед найбільш вивчених розрядів, таких як антропоніми та топоніми, дослідники не мають єдиної думки щодо класифікації.

По-друге, головним критерієм включення до корпусу буде визначення власної назви, що й стане одиницею зберігання. Власна назва — індивідуальне найменування окремих одиничних об'єктів [3, с. 83]. Послугуючись таким визначенням, ми будемо включати у корпус широкий спектр власних назв як вигаданих, на кшталт *Перевернипляшка*, так і цілком реальних осіб та явищ у будь-якому дискурсі.

З цього виходить третій аспект розмітки Корпусу, який здається нам найбільш вдалим: розмітка за жанром: поетичний, розмовний, електронний, газетний, розмовний тощо. Звичайно, існують такі проблеми для складання корпусу, як обмеження у використанні програмного забезпечення, авторських прав та питання етики, доступність текстів взагалі, коли ми говоримо про історичні тексти або рукописи.

Проект Ономастичного корпусу планується для широкого кола відвідувачів та дослідників мови. Корпус створюватиметься не тільки для українських дослідників з лінгвістики, а й для міжнародного наукового товариства. Може статися в нагоді й науковцям у таких галузях як культура, філософія, соціологія та історія. Корпус значно полегшить працю над словниками мови письменників, загальними та культурно-філософськими. Ономастичний корпус надає можливість для міжкультурних, компаративістських студій, його різноаспектне та багаторазове використання відкриває горизонт для створення нових лінгвістичних продуктів.

### Список використаної літератури

1. Демська О. М. Теорія корпусної параметризації сучасної української мови : автореф. дис. ... д-ра філол. наук / Демська Оріся Мар'янівна. — Одеса, 2012. — 29 с.
2. Дмитриев А. В. Корпус источников в автоматизированной топонимической базе данных Ингерманландии / А. В. Дмитриев // Труды международной научной конференции «Корпусная лингвистика-2011». — М., 2011. — С. 163- 168.
3. Карпенко Ю. О. Власні назви / О. Ю. Карпенко // Українська мова : Енциклопедія. — 2-ге вид, випр. і доп. — К.: Вид-во «Українська енциклопедія» ім. М. П. Бажана, 2004. — С. 83- 84.
4. Перцов Н. В. О роли корпусов в лингвистических исследованиях / Н. В. Перцов // Труды международной конференции «Корпусная лингвистика-2006». — М., 2006. — С. 318-331.
5. Плунгян В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики / Плунгян В. А. // Русский язык в научном освещении. 2 (16). — 2008. — С. 7-20.
6. Рахилина Е. В. Корпус как творческий проект / Е. В. Рахилина // Национальный корпус русского языка: 2006- 2008. — 2009. — С. 7-26.
7. Резникова Т. И. Славянская корпусная лингвистика / Т. И. Резникова // Die Welt der Slaven LIII. — 2008. — P. 10-38.
8. Рыков В. В. Прагматически ориентированный корпус текстов / Рыков В. В. — Режим доступа: <http://rykov-cl.narod.ru/t.html>
9. Baker P. A Glossary of Corpus Linguistics / Baker P. — Edinburgh: Edinburgh University Press Ltd., 2006. — 188 p.
10. Meyer Ch. English Corpus Linguistics: An Introduction / Meyer Charles. — Cambridge: Cambridge University Press, 2002. — 184 p.
11. Xiao R. Corpus Creation / R. Xiao // Handbook of Natural Language Processing. — 2<sup>nd</sup> ed. — London, 2010. — P. 147- 166.

12. <http://www.anc.org/>
13. <http://corpus.byu.edu/bnc/>
14. <http://khnt.aksis.uib.no/icame/manuals/brown/>
15. <http://www.natcorp.ox.ac.uk/>
16. <http://www.mova.info/corpus.aspx?11=209>
17. <http://tatoeba.org>
18. <http://toris.krc.karelia.ru/info/news.ru.phtml>
19. TIME Magazine Corpus

**Неклесова В. Ю.**

*кандидат филологических наук, доцент кафедры иностранных языков гуманитарных факультетов*

*Одесского национального университета им. П. П. Мечникова e-mail:*

[neklesova.valerie@gmail.com](mailto:neklesova.valerie@gmail.com)

**ФОРМИРОВАНИЕ УКРАИНСКОГО ОНОМАСТИЧЕСКОГО КОРПУСА**

**Резюме**

*В статье рассматривается вопрос образования и компиляции украинского ономастического корпуса. Изучены уже существующие корпуса, а также те корпуса, которые имеют данные украинского языка.*

**Ключевые слова:** ономастика, украинский ономастический корпус.

**Neklesova Valeriia**

*PhD, Associate Professor,*

*Chair of Foreign Languages of the Arts Faculties Odessa I.I. Mechnikov National University e-*

*mail: [neklesova.valerie@gmail.com](mailto:neklesova.valerie@gmail.com)*

**FORMING UKRAINIAN ONOMASTIC CORPUS**

**Abstract**

*The article dwells on the problems of development of the Ukrainian Onomastic Corpus. The aim of the article is to inform readers with the state of corpora in Ukrainian linguistic studies. The National Project of Ukrainian Corpus has been gradually growing for more than ten years already but still the Ukrainian corpus linguistics is at the beginning of its way.*

*The project of the Ukrainian Onomastic Corpus is discussed in this article. We propose to take into consideration and as an example already existing corpora such as The Brown University Standard Corpus of Present-Day American English (or just Brown Corpus), The British National Corpus (BNC), The American National Corpus (ANC).*

*As becomes clear from its name the key words for the Corpus are proper names. The annotation of the corpus is planned in different ways: time aspect (historical texts and contemporary), lexical level (lemmata), stylistic features (fiction, documentary, academic, e-communication, formal, informal, etc.)*

**Key words:** *onomastics, Ukrainian Onomastic Corpus.*

*Надійшла до редакції 23.06.2015*