

Hlyvins'ka, Lesja Kostjantynivna

**Корпус української мови як джерело лексикографічної рецепції
художнього тексту**

Opera Slavica. 2013, vol. 23, iss. 4, pp. 43-50

ISSN 1211-7676 (print); ISSN 2336-4459 (online)

Stable URL (handle): <https://hdl.handle.net/11222.digilib/129001>

Access Date: 27. 11. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

КОРПУС УКРАЇНСЬКОЇ МОВИ ЯК ДЖЕРЕЛО ЛЕКСИКОГРАФІЧНОЇ РЕЦЕПЦІЇ ХУДОЖНЬОГО ТЕКСТУ

Леся Глівінська (Київ)

Абстракт:

Корпусна лінгвістика пропонує істотну підтримку всім мовознавчим галузям, у тому числі й авторській лексикографії. Потенціал *Корпусу української мови* можна ефективно використовувати в аспекті параметризації художніх текстів, зокрема: формувати дослідницьку базу даних на основі «робочого» підкорпусу (за творами письменника); структурувати відомості про будь-який рівень мовної ідіосистеми – фонетичний, морфемний, лексичний, синтаксичний; діагностувати лінгвальні особливості авторського тексту, враховуючи жанрову, тематичну, хронологічну специфіку обстежуваного континууму; проектувати на обраному матеріалі різноманітні словники (частотні, ідеографічні, конкордансного типу).

Ключові слова: Корпус української мови, авторська лексикографія, художній текст

Ukrainian Corpus as a Source of Lexicographic Reception by Literary Text

Abstract:

Corpus linguistics suggests substantial support to all linguistic branches including author's lexicography. Potential of the *Ukrainian Corpus* may be used effectively in relation to parameterization of literary texts, in particular to form a research data base using "work" subcorpus (by writer's works); to structure information of any level of linguistic idiosystem, namely phonetic, morphemic, lexical, syntactic; to diagnose lingual peculiarities of the author's text considering genre, thematic, chronological specific character of the continuum under examination; based on the selected material to design various dictionaries (frequency, ideographic, concordance).

Key words: Ukrainian Corpus, author's lexicography, literary text

На сучасному етапі у світовій лінгвістиці спостерігається надзвичайне піднесення лексикографічної активності. З розвитком комп'ютерних технологій зникли обмеження на обсяг аналізованого матеріалу – з'явилася можливість продуктивно працювати з надвеликими масивами текстів. Мова постала в досі незміряному розмаїтті фактів. Це закономірно стимулювало фахівців до оптимізації словникової роботи.

Питання лексикографії залишаються пріоритетними й на тлі *українознавчої* проблематики¹. Водночас маємо констатувати, що вітчизняний досвід *авторського* словникарства потребує поглибленого вивчення в контексті реалій та перспектив наукового пошуку. (Пор.: «Хоча творення АС² не належить до інноваційного розділу української лексикографії, все ж чимало загальних і часткових питань... не зауважено критикою, залишено без належного аналізу теоретиками-лексикографами. Досі здебільшого в загальних рисах визначають роль АС для лінгвістики, літературознавства, зокрема текстології...» (Гриценко, 2012, 41).)

Уже традиційно авторську лексикографію кваліфікують як теорію і практику укладання *письменницьких* словників. Ідіолекти поетів, прозаїків, драматургів дістають поліаспектної параметризації в різножанрових лексиконах (конкордансах, частотних словниках, глосаріях і под.).

Важливо, що істотну підтримку всім мовознавчим галузям, у тому числі й авторській лексикографії, пропонує нині *корпусна лінгвістика*. Електронний гіпертекстовий ресурс, призначений для наукового вивчення мови, – тобто корпус, – справедливо називають «центром всесвіту словникарства» (Демська, 2011b, 34)³ і водночас визнають одним із найбільш доцільних способів інтерпретації авторського, зокрема художнього, мовлення (Шестакова, 2012, 18–19).

Ліностративний чи дослідницький; моніторинговий чи референційний; динамічний або статичний⁴ – незалежно від мети і призначення, формату текстового матеріалу і типу програмної обробки, – корпус зберігає свою основну перевагу, тобто репрезентує мову такою, якою вона «є насправді, а не якою ми хочемо, щоб вона була» (Плунгян). Зрештою, «мова як предмет лінгвістики і текст як матеріальний вияв мови, придатний для наукових студій, не може бути... фантазією вченого» (Демська, 2011a, 137). І саме робота з вичерпними корпусними даними – це гарантія достовірного, об'єктивного результату.

В українській науці спостерігаємо поступальне утвердження корпусних стратегій. У провідних дослідницьких закладах (Національна академія наук України, Київський національний університет імені Тараса Шевченка, Львівський національний університет імені Івана Франка, Волинський національний університет імені Лесі Українки та ін.) триває підготовка загальномовного і спеціалізованих корпусів. Щоправда, доступними для онлайн-користування є лише окремі ресурси, з-поміж яких – і ті, що відображають ідіолектний зріз національної мови. Так, об'єктом корпусної рецепції стала проза Івана Франка⁵ (КФ); різножанрова спадщина

¹ Див., напр., новітні спеціалізовані видання: *Українська лексикографія в загальнослов'янському контексті: теорія, практика, типологія*. Київ, 2011; *Українська і слов'янська тлумачна та перекладна лексикографія*. Київ, 2012.

² Авторських словників.

³ Доречно згадати, що в тематичному збірнику «Проблеми і методи лексикографії» свого часу було опубліковано програмне дослідження В. Н. Френсіса – одного з основоположників світової корпусної лінгвістики (Френсіс, 1983).

⁴ Докладніше про типологію корпусів ідеться в (Демська, 2011a, 89–101; Карпіловська, 2006, 74–79).

⁵ Іван Якович Франко (1866–1916) – український поет, прозаїк, драматург, перекладач, лінгвіст, літературознавець, текстолог, фольклорист, етнограф, історик, театрознавець,

письменників XIX – XXI ст. (як джерельна база *Корпусу української мови* (КорУМ)); сучасна українська поезія (ПБД). Авторським корпусом доцільно вважати конкорданс творів Григорія Сковороди⁶ (ОСон), спроектований зарубіжними й вітчизняними фахівцями.

У контексті заявленої статті розглянемо функціональні переваги *Корпусу української мови* (КорУМ)⁷, що можуть бути актуальними для дослідника естетичної комунікації й зокрема, для упорядника авторського словника.

Відзначимо, що інформативність корпусу безпосередньо корелює із прогнозованими сферами його використання. При цьому корпуси загальнонаціонального статусу мають якнайповніше відбивати стильову стратифікацію етнічної мови. І важливо, що основою таких електронних ресурсів є тексти *художнього* стилю. Ідеться навіть про певну канонізацію творів авторитетного письменника як об'єктів лінгвістичного обстеження (Демська, 2011а, 49).

У *Корпусі української мови*, генеральний масив якого становить нині близько 20 млн слововживань⁸, 40 % (майже 8 млн сл.) складають тексти *художньої* літератури, а саме: прозові твори – до 35 % (до 7 млн сл.); поетичні – приблизно 5 % (майже 1 млн сл.). Список джерел містить імена понад 100 творців нової та сучасної української художньої словесності (XIX–XXI ст.). Зауважимо, що корпус є динамічним, відтак передбачає перманентне поповнення множини текстів. А отже, список літературів і відповідно їхніх творів-об'єктів корпусної інтерпретації залишається відкритим.

Відомо, що проблема систематики вихідного лінгвального матеріалу з особливою гостротою постає перед спеціалістом-лексикографом. Але за наявності корпусу ця проблема зникає. Електронний текстовий ресурс загальномовного типу дозволяє швидко і якісно сформувати дослідницьку базу даних на основі не лише генерального корпусного масиву, а й «робочого» підкорпусу, виділеного, наприклад, за творами певного письменника. Окресливши мету наукового пошуку і гіпотетичний шлях

публіцист, громадсько-політичний діяч. В Україні започатковано підготовку академічного зібрання його творів у 100 томах.

⁶ Григорій Савич Сковорода (1722–1794) – найвидатніша постать у духовно-культурному житті України XVIII ст. Поет, байкар, перекладач, теоретик літератури, філософ, мандрівний педагог-просвітитель.

⁷ На сьогодні в нашій державі реалізується кілька незалежних проектів зі створення загальномовного корпусу; див., напр.: (Широков, 2011; Демська, 2011а, 155–186). Однак єдиним ресурсом відкритого типу залишається дослідницький *Корпус української мови* (тестова версія), над яким працюють учені лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка.

⁸ За обсягом аналізований корпус належить до розряду середніх. Вважається, що фундаментальні корпуси мають налічувати щонайменше 100 млн слововживань. Водночас т. зв. мегакорпуси вже містять понад 700 млн сл. (див. докладніше: (Демська, 2011а, 100–101, 130)). Представницькі електронні текстові ресурси створено для багатьох мов, зокрема слов'янських: російської, чеської, польської, словацької, болгарської, македонської, словенської, сербської, хорватської. Показово, що в межах однієї лінгвістичної традиції розбудовуються корпуси різних типів і призначення.

її реалізації, лінгвіст може впорядкувати різноаспектну інформацію про будь-який рівень мовної ідіосистеми – фонетичний, морфемний, лексичний, синтаксичний; при цьому діагностувати особливості функціонування мови в художньому тексті, враховуючи жанрову, тематичну, хронологічну специфіку обстежуваного континууму. Відтак показовою є роль корпусного ресурсу як *інтелектуальної бази даних* – джерела знань про індивідуально-авторську мовну *картину світу*.

Підкреслимо, що максимальну дієвість *Корпусу української мови* має забезпечити його *автоматична* лінгвістична розмітка⁹. Такій раціональній стратегії надали перевагу розробники проекту¹⁰. На сьогодні здійснено *морфологічне анотування* корпусного матеріалу, яке передбачає характеристику слова в аспекті частиномовної належності й категоріальних значень (роду, числа, відмінка, особи тощо). Варто підкреслити, що ефективність автоматичного морфологічного аналізу українських текстів становить близько 96 % (Дарчук, 2012, 244), отже, є вельми показовою. Зрештою, маємо вагомий інструментарій для цілісного аналізу парадигматики авторської мови; для розкриття функціонального потенціалу одиниць певних частиномовних класів; для вивчення граматичних категорій щодо їх використання в стилістичному плані; для оцінки ступеня вживання словоформ (за критеріями загальної поширеності, рівномірності розподілу, стабільності); для опису морфологічних синонімів як засобу стилістики тощо. Дослідницьку роботу ймовірно скерувати в гендерному напрямку: це дозволяють пошукові можливості *Корпусу української мови*. Адже кореляції в лексико-граматичних ідіосистемах – потенційний чинник гендерної варіативності мовної (ширше – концептуальної) картини світу.

Зауважимо, що нині триває робота з *морфемного індексування* корпусу, яке полягає в сегментації текстів на морфи. На цьому етапі розмітка здійснюється за методикою комп'ютерного моделювання морфемних структур українських слів. Суттєво, що програмне забезпечення уможливує не тільки якісний аналіз словоформ, а й проектування морфемних і словотвірних картотек / словників в автоматизованому чи автоматичному режимі. Об'єктом опису за вказаною методикою став поки що фрагмент корпусного матеріалу, а саме – вибірки з художніх текстів Тараса Шевченка¹¹, Лесі Українки¹², Василя Стуса¹³ і Ліни Костенко¹⁴. Авторські частотні

⁹ Терміни «розмітка», «анотування», «маркування», «індексування» використовують як взаємозамінні, коментуючи процедуру формалізації корпусних даних.

¹⁰ Разом із тим фахівці визнають, що маркування корпусу самим дослідником (за допомогою спеціальних комп'ютерних програм) детермінує вірогідніші результати (Дарчук, 2012, 238–239). Зрештою, досягти абсолютної точності при автоматичному аналізі тексту практично неможливо.

¹¹ Тарас Григорович Шевченко (1814–1861) – вершинна постать в історії української духовної культури. Поет, прозаїк, драматург, художник, мислитель; основоположник нової української літератури і нової української літературної мови.

¹² Леся Українка (справжнє ім'я – Лариса Петрівна Косач-Квітка; 1871–1913) – поетеса, прозаїк, драматург, перекладач, літературний критик, фольклорист, публіцист, громадська діячка.

¹³ Василь Семенович Стус (1938–1985) – український поет-дисидент, прозаїк, перекладач, літературознавець, правозахисник.

словники морфем і морфемних структур викладено на сторінках лінгвістичного порталу MOVA.info [<http://www.mova.info/article.aspx?l1=210&DID=5215>] 17.04.2013].

Вдосконаленню лінгвістичної бази *Корпусу української мови* має сприяти й новітній лексикографічний продукт – тлумачний словник афіксальних морфем, над створенням якого працюють вітчизняні фахівці (Дарчук, 2012, 243). Такий ресурс стане надійним підґрунтям для вивчення комбінаторно-дистрибутивної будови авторського слова і ширше – для пізнання індивідуальних закономірностей лінгво-креативної практики.

Завданням на перспективу для упорядників *Корпусу української мови* залишається вироблення детальних принципів розмітки за фонетичними, семантичними і синтаксичними параметрами.

Так, автоматизація процесів транскрибування і транслітерування здатна істотно спростити функції лінгвіста. Зауважимо, що *фонетичну розмітку* доцільно передбачити для текстів, зокрема літературних, що мають виразні фонемографічні й орфографічні особливості. Відтак узагальнені відомості будуть релевантними в описі мовного «портрета» митця чи епохи, у яку він жив або яку художньо відтворив. На підставі фонетичного маркування можна вивчати звукопис, авторську просодіку, метричну організацію текстів.

Беззаперечні переваги матиме й *семантична розмітка* корпусу, орієнтована на вивчення слова в лексикологічному аспекті. Спеціальні маркери можуть вказати на сегменти ключові й другорядні в авторському словнику – зі складу лексики активної/пасивної; стилістично забарвленої/нейтральної; питомої/запозиченої. Переваги текстового корпусу доцільно використати для аналізу ідеографічних зв'язків між різночастининомовними одиницями. Водночас актуалізація таких системних відношень, як полісемія, омонімія, синонімія, паронімія, антонімія, скеровує увагу дослідника на тенденції слововживання – загальномовні та індивідуально-авторські. Матеріал конкордансних ілюстрацій унаочнює закономірності смислової динаміки, зреалізовані в тексті.

Повноті аналізу авторської мови неодмінно сприятиме *синтаксичне маркування*. На цьому етапі передбачено здійснити виділення словосполучень і їх структурно-семантичну кваліфікацію, а також побудову дерева залежностей для речень. За наявності відповідного інструментарію можна ефективно досліджувати валентність українського слова, формозмістові параметри синтаксичних одиниць і зрештою, дериваційні зв'язки між різними сегментами мовної структури.

Доцільно згадати й про інші способи лінгвістичного анутовання корпусного матеріалу. Якісно нові відомості про естетичну комунікацію можна здобути, наприклад, на підставі *анафоричної розмітки*, тобто визначення кореферентних відношень між елементами висловлювання. Цей тип анотації використовують, зокрема, укладачі *Корпусу текстів Івана Франка* (Бук, 2010). Як правило, анафоричними¹⁵ є займенникові слова, що в цілому покривають значний відсоток мовно-художнього

¹⁴ Ліна Василівна Костенко (нар. 1930 р.) – українська письменниця, публіцист, громадська діячка.

¹⁵ *Анафоричними* (від грецьк. ἀναφορά, букв. – піднесення) називають слова й словосполучення, що відсилають до попереднього контексту, до раніше сказаного (Ахманова, 2004, 47).

континууму¹⁶, а тому виявлення їх зв'язку з номінаціями конкретних об'єктів дійсності дозволяє повніше представити частотно-тематичний спектр аналізованих текстів. Принагідно зауважимо, що один із важливих етапів корпусного «портретування» Франкової мови передбачає *ономастичну індексацію* словника: кожна власна назва супроводжується вказівкою на її тип (ім'я, прізвище, місто, країна і под.). Такий підхід до аналізу ідіолекту вже зреалізовано на матеріалі великої прози письменника; див., напр.: (Бук, 2012). Очевидно, що наявність ономастичних маркерів не тільки уможливило якісну підготовку емпіричної бази, а й оптимізує подальшу інтерпретацію власних назв як концептуальних одиниць авторського тексту.

Корисною для дослідника поезії могла б стати *метрична розмітка* корпусу, що дозволила б ефективно вивчати віршове мовлення в аспекті його формальних особливостей – метрики, ритмики, строфіки тощо. Прикладом такого ресурсу є *Поетичний корпус* (ПоетКор), створений російськими вченими. Розробники проекту переконані, до речі, що корпусна інтерпретація поезії принципово відрізняється від аналізу прози (художньої та нехудожньої). Тому розмітку поетичних і прозаїчних текстів доцільно здійснювати не на загальних підставах, а диференційовано. Віршова мова потребує особливого індексування (Гришина, 72–73).

Зауважимо, що формалізація лінгвістичних відомостей у корпусі може мати характер *дискурсної анотації*, коли об'єктом маркування стають системові елементи дискурсу. Щоправда, цей спосіб розмітки не набув поширення «через неоднозначність ідентифікації дискурсної інформації в тексті» (Демська-Кульчицька, 2003, 41).

Варіант *прагматичного індексування* також прийнятний для корпусної побудови (Демська, 2011а, 187). Йдеться про експлікацію тих мовних фактів, появу яких можна пояснити впливом конкретних екстралінгвальних чинників (історичних, соціокультурних, етнопсихологічних та ін.).

Отже, підходи до маркування електронного текстового корпусу бувають різні. Їх вибір залежить від ракурсу дослідницького пошуку. Найширші можливості для наукової рефлексії надають корпуси з розгалуженою системою параметризації.

Закономірно, що інноваційні процеси в сучасній лексикографії пов'язані з утвердженням *корпусних* стратегій. Зрештою, корпус і є тим «живим простором», у якому функціонує мова (Плунгян). Вичерпний і репрезентативний електронний текстовий ресурс «робить можливим значне розширення емпіричної лексикографічної бази і, відповідно, створює умови для розбудови реєстрів словників, пропонує значно більший ілюстративний матеріал, стимулює появу нових типів лексикографічного опису, нових типів словників» (Демська, 2010, 171). Підкреслимо, що «багатопараметровість» – фундаментальна перевага саме інтерактивних лінгвістичних ресурсів, адже користувач генерує лише ту інформацію, яка безпосередньо пов'язана з проблематикою дослідження. (Що ж до традиційних джерел – «паперових» словників, – які мають обмежений обсяг і фактично є закритою інформаційною системою, то наявність усіх можливих характеристик вокабули «ще не вказує на більшу ефективність лексикографічної праці» (Балог, 2004, 39–40).)

¹⁶ Пор. розподіл лексико-граматичних класів, наприклад, в електронних словниках сучасної української поезії (ЧСа) та прози (ЧСб).

Важливо, що *Корпус української мови*, який нині активно розбудовується, є поліфункціональним ресурсом. Його інструментарій уже сьогодні дозволяє працювати над комплексом масштабних теоретико-прикладних завдань. Без сумніву, «лише докладне лексикологічне обстеження великої кількості різностильових текстів, що творять сукупний сучасний текстовий простір українського літературного ідіому, ...здатне створити якісно нову інформаційну ситуацію, живильну для наступного розвитку і української мови, і лінгвістики в цілому» (Гриценко, 2011, 248–249).

Маємо підстави стверджувати, що *Корпус української мови* є надійним підґрунтям у виборі вихідних принципів лексикографічного аналізу *художніх* текстів. Самі реалії сучасного культурного розвитку з його антропоцентричною спрямованістю детермінують потребу в «дієвих» *авторських* словниках як таких, що всебічно моделюють результати мовної практики творчої особистості. Потенціал *Корпусу української мови* можна ефективно використовувати в різноманітних аспектах, зокрема таких:

- порівнювати генеральну текстову базу й авторські підкорпуси з метою виявлення динамічних процесів, аналізу ядерних/периферійних явищ, встановлення типових/індивідуальних фактів;

- інтегрувати або диференціювати відомості фонетичного, граматичного, семантичного, стилістичного, комунікативно-прагматичного характеру;

- проектувати на обраному матеріалі різноманітні словники (частотні¹⁷, ідеографічні, конкордансного типу та ін.).

Отже, *Корпус української мови* – помітне досягнення вітчизняної науки. Створений на ідейних засадах сучасної лінгвістики і з урахуванням новітніх можливостей інформаційно-комп'ютерного середовища, він є релевантним дослідницьким зрядом для гуманітаріїв різного профілю.

Література:

- АХМАНОВА, О. С.: *Словарь лингвистических терминов*. Москва, КомКнига, 2004.
- БАЛОГ, В.: *Параметрична система тлумачних словників української мови та її використання в комп'ютерній лексикографії*. Ін: Лексикографічний бюлетень, 2004, вип. 10, 25–40.
- БУК, С. Н.: *Розрізнення анафори у корпусі текстів Івана Франка: «я» у «Перехресних стежках»*. Ін: Вісник Луганського національного університету імені Тараса Шевченка, 2010, № 11 (ч. II), 151–162.
- БУК, С. Н.: *Частотний словник роману Івана Франка «Основи суспільності». Інтерпретація твору крізь призму статистичної лексикографії*. Львів, ЛНУ ім. Івана Франка, 2012.

¹⁷ Алфавітно-частотні реєстри лексем і словоформ можна впорядкувати в *автоматичному* режимі до будь-якого корпусного тексту. Для одиниць опису визначено базову квантитативну характеристику – абсолютну частоту. Надалі статистичний «портрет» художньої мови мають доповнити ще три показники: середня частота, міра колювання середньої частоти (середнє квадратичне відхилення), коефіцієнт стабільності.

- ГРИЦЕНКО, П. Ю.: *Авторська лексикографія: Тарас Шевченко*. In: Українська і слов'янська тлумачна та перекладна лексикографія. Київ 2012, 40–69.
- ГРИШИНА, Е. А. И ДР.: *Поэтический корпус в рамках Национального корпуса русского языка: общая структура и перспективы использования*. In: <http://ruscorpora.ru/sbornik2008/05.pdf> (29. 7. 2013).
- ДАРЧУК, Н. П. ТА ІН.: *Корпус української мови як джерело для мовознавчих досліджень*. In: Українська і слов'янська тлумачна та перекладна лексикографія. Київ 2012, 238–245.
- ДЕМСЬКА, О.: *Текстовий корпус: ідея іншої форми*. Київ, ВПЦ НаУКМА, 2011а.
- ДЕМСЬКА, О.: *Українська корпусна лінгвістика: сьогодення і перспективи*. In: Дивослово, 2011b, № 10, 32–34.
- ДЕМСЬКА, О. М.: *Вступ до лексикографії*. Київ, Києво-Могилянська академія, 2010.
- ДЕМСЬКА-КУЛЬЧИЦЬКА, О. М.: *Базові поняття корпусної лінгвістики*. In: Українська мова, 2003, № 1, 40–45.
- КАРПІЛОВСЬКА, Є. А.: *Вступ до прикладної лінгвістики: комп'ютерна лінгвістика*. Донецьк, Юго-Восток, 2006.
- КорУМ: *Корпус української мови*. In: <http://www.mova.info/corpus.aspx?11=209> (28. 5. 2013).
- КФ: Бук, С. та ін. *Онлайн-конкорданс роману Івана Франка «Перехресні стежки»*. In: <http://www.ktf.franko.lviv.ua/~andrij/science/Franko/concordance.html> (17. 4. 2013).
- ПБД: *Параметризована база даних сучасної української поетичної лексики*. In: <http://www.mova.info/page3.aspx?11=96> (28. 5. 2013).
- ПЛУНГЯН, В. А.: *Почему современная лингвистика должна быть лингвистикой корпусов*. In: <http://polit.ru/article/2009/10/23/corpus/> (17. 4. 2013).
- ПоетКор: *Поэтический корпус*. In: <http://www.ruscorpora.ru/search-poetic.html> (2. 7. 2013).
- ФРЕНСИС, У. Н.: *Проблемы формирования и машинного представления большого корпуса текстов*. In: Новое в зарубежной лингвистике, 1983, вып. XIV «Проблемы и методы лексикографии», 334–368.
- ЧСа: *Частотний словник сучасної української поетичної мови. Загальний словник граматичних класів*. In: <http://www.mova.info/poetgr.aspx?poet=poets> (28. 5. 2013).
- ЧСь: *Частотний словник сучасної української прози*. In: <http://www.mova.info/Page2.aspx?11=90> (17. 4. 2013).
- ШЕСТАКОВА, Л. Л.: *Русская авторская лексикография: теория, история, современность*. Автореф. дис. д-ра филол. наук. Москва, 2012.
- ШИРОКОВ, В. А. ТА ІН.: *Застосування Українського національного лінгвістичного корпусу в лексикографії та лінгвістичних експертизах*. In: Українська лексикографія в загальнослов'янському контексті: теорія, практика, типологія. Київ 2011, 285–294.
- ОСon: *Online Concordance to the Complete Works of Hryhorii Skovoroda*. Project created by Natalia Pylypiuk and Oleh S. Hnytzkyj. In: <http://www.artsrn.ualberta.ca/ukr/skovoroda/NEW/index.php?glang=> (17. 4. 2013).