

Отримано: 7 березня 2018 р.

Прорецензовано: 13 березня 2018 р.

Прийнято до друку: 15 березня 2018 р.

e-mail: helen.tkachyk@gmail.com

DOI: 10.25264/2519-2558-2018-1(69)/2-194-196

Ткачик О. В. Використання корпусних методів у гендерних дослідженнях. *Наукові записки Національного університету «Острозька академія»: серія «Філологія»*. Острог: Вид-во НаУОА, 2018. Вип. 1(69), ч. 2, березень. С. 194–196.

УДК 81'33-114

Ткачик Олена Володимирівна,
Київський політехнічний інститут ім. Ігоря Сікорського, м. Київ

ВИКОРИСТАННЯ КОРПУСНИХ МЕТОДІВ У ГЕНДЕРНИХ ДОСЛІДЖЕННЯХ

У статті розглянуто способи використання методів корпусної лінгвістики для дослідження гендеру в мові на прикладах останніх наукових розвідок у галузі гендерної лінгвістики, заснованих на обробках корпусів. Наведено приклади доповнення різних аспектів сучасних гендерних досліджень, а саме культурологічного, соціолінгвістичного, дискурсивного, лінгвокогнітивного та ін. якісно та кількісно новими результатами. Кількісний та якісний аналізи великих масивів даних можливий завдяки таким методам, як укладання списків частотності слововживань, списків колокатів, аналіз конкордансу, аналіз ключових слів, семантичне тегування, використання референтних корпусів тощо.

Ключові слова: корпус, корпусні методи, корпусна лінгвістика, гендер, чоловік, жінка, фемінність, маскуліність.

Ткачик Елена Владимировна,
Киевский политехнический институт им. Игоря Сикорского, г. Киев

ИСПОЛЬЗОВАНИЕ КОРПУСНЫХ МЕТОДОВ В ГЕНДЕРНЫХ ИССЛЕДОВАНИЯХ

В статье рассматриваются способы использования методов корпусной лингвистики для исследования гендера в языке на примерах анализа последних научных изысканий в области гендерной лингвистики, основанных на обработках корпусных данных. Количественный и качественный анализ огромных массивов данных возможен благодаря таким методам, как составление списков частотных словоупотреблений, списков колокатов, анализ конкорданса, анализ ключевых слов, семантическое тегирование, использование референтных корпусов и т.д.

Ключевые слова: корпус, корпусные методы, корпусная лингвистика, гендер, мужчина, женщина, феминность, маскулинность.

Olena Tkachyk,
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv

USING CORPUS METHODS TO ANALYZE GENDER

The article is intended to review the application of corpus linguistics methods to the study of gender by examining examples of recent investigations in the sphere of gender linguistics which represent recent methodological advances of corpus methods. The seven selected works, used to demonstrate this, apply a range of techniques that corpus linguistics can offer researchers in gender analysis. The authors of mentioned papers assemble their own corpora and address the most appropriate techniques to reach the goals set in their research. The methods used are compilation of frequency lists, concordance analysis, collocates, keyness, semantic tagging and employment of reference corpora, which allows a combination of qualitative and quantitative approaches to the study. Application of corpus methods in the investigation of gender and language demonstrates among others a number of advantages: operating large collections of texts, quick and reliable search, receiving statistical data, possibility to mix automated and hand approaches, obtaining results which are free of ideological biases and objectivity of conclusions. This article exemplifies and summarizes how cultural, social, psychological, discursive, cognitive and other approaches to the study of gender in the language are accomplished with new qualitative and quantitative results.

Key words: corpus, corpus linguistics, gender, man, woman, feminine, masculine.

Традиційна гендерна лінгвістика репрезентована різними напрямками досліджень: культурологічними розвідками гендерного концептуального змісту на матеріалі лексики (О.Л.Бессонова, А.В.Дудолодова, Г.І.Емірсуїнова, М.С.Колесникова, А.П.Мартинюк, І.Г.Ольшанський, Г.Ю.Цапро), фонетики (Т.Н.Шевченко), фразеології (І.В.Зикова, Т.А.Клименкова, О.П.Левченко, Д.Малишевська), пареміології (Ю.В.Абрамова, О.Ю.Дубенко, А.В.Кириліна); також соціолінгвістичними (В.В.Потапов), психолінгвістичними (О.А.Андріанова, О.І.Горошко) та дискурсивними (О.Л.Антинескул, Г.Ю.Богданович, Л.Ф.Компанцева, Е.В.Піщікова, Л.М.Синельникова, І.І.Халеева) дослідженнями мовленнєвої поведінки статей; аналізом гендерного аспекту художнього дискурсу (Н.Д.Борисенко, О.Л.Козачишина, І.І.Морозова). Одним із напрямів гендерних досліджень є порівняльний аналіз гендерної комунікативної поведінки різних вікових груп (Е.В.Піщікова), професійних груп (О.І.Горошко), аналіз статусів персонажів (Н.Д.Борисенко, В.А.Маслова), а також невербальних засобів комунікації (Л.В.Солошук, Л.І.Коваленко). Дискурсивна онтологія дає нове розуміння гендеру як соціокультурної категорії, конструйованої у процесі інтерактивної взаємодії індивідів у соціальному контексті (D.Cameron, J.Coates).

Комп'ютерна епоха надає нові можливості для сучасних гендерних досліджень, додаючи до культурологічного, соціолінгвістичного, психолінгвістичного, лінгвокогнітивного та ін. наукових підходів якісно та кількісно нові результати. У західних дослідженнях гендеру та мови корпусні методи є досить популярними, але для вітчизняної лінгвістики вони ще доволі нові та мало застосовані. У статті аналізуються останні розвідки, які представляють собою недавні методологічні досягнення у галузі гендерної лінгвістики і демонструють використання підходів корпусного аналізу при дослідженні питань, що стосуються гендеру та мови.

McENERY та HARDIE визначають лінгвістику корпусів як «групу методів вивчення мови, що застосовуються до деякого набору машиночитаних текстів, або корпусу, який зазвичай має розмір, непідвласний аналізу вручну в межах будь-яких часових рамок... Корпусом неодмінно користуються за допомогою інструментів, які дозволяють швидкий та надійний пошук» [5, с.1-2].

Сучасні українські гендерні дослідження з використанням методів корпусної лінгвістики нечисленні і проводяться, зокрема, проф. О.П.Левченко, яка відмічає, що «використання значних масивів даних є важливою умовою отримання релевантних результатів» [1]. Базуючись на дослідницькому корпусі з уривків художньої прози українських письменниць та письменників, авторка виявляє типові порівняльні стратегії, наявні в сучасному українському художньому дискурсі. Специфіка вживання засобів порівняння в мовленні жінок та чоловіків розглядаються як способи «компаративної» вербалізації певних концептосфер авторів-чоловіків і жінок.

За допомогою аналізу порівнянь отримуються висновки про атрибути, які належать тому чи іншому концепту в картинах світу жінок та чоловіків, що, в свою чергу, стає підґрунтям для аналізу типових уявлень носіїв сучасної української мови про маскуліність і фемініність.

Дослідницький корпус, який став підґрунтям цієї праці, містить підкорпуси «чоловічої» і «жіночої» прози, що налічують приблизно по 460 тис. слововживань кожний. За допомогою програми AntConc3.2.4w створено пілотні конкорданси для сполучників *як, наче, неначе, мов, немов, мовби*. Ручним сортуванням сформовано переліки порівнянь та порівняльних зворотів: у текстах авторок-жінок налічується 1860 одиниць, в авторів-чоловіків 996 одиниць. Кількісні дані засвідчують виразну асиметрію в уживанні порівнянь жінками й чоловіками (у жіночих текстах порівнянь на 46% відсотків більше, ніж у чоловічих). Виявлено, що і жінки і чоловіки використовують порівняння для вербалізації виділених концептосфер у практично однаковому співвідношенні, проте терміни на позначення концептосфери Тварина у «жіночих» текстах зустрічається 179 разів, натомість у «чоловічих» – 106. Застосувавши корпусний аналіз, О.П.Левченко приходять до висновку, що чоловіки частіше використовують стійкі порівняння з певним компонентом, а у «жіночих» текстах більше індивідуально-авторських одиниць [1].

Звертаючись до досвіду використання корпусних методів для дослідження гендеру в мові західними лінгвістами, зазначаємо роботу Пола Бейкера «Introduction: Virtual Special Issue of Gender and Language on corpus approaches» [3], в якій він робить огляд таких джерел: Пол Бейкер, Джонсон і Енслін (2007), Хартіс-Чорний та Сіле (2009), Холмгрен (2009), Бейкер (2010) та Кінг (2011), де автори використовували методіку корпусної лінгвістики в своїх дослідженнях, поєднуючи якісні та кількісні підходи.

Джонсон і Енслін (2007) побудували корпус британських газетних статей для вивчення гендерних ідеологій у мові. Фокусуючись на двох термінах «її-мова» та «його-мова», вони змогли порівняти контексти, в яких згадувались чоловіки та жінки щодо їх використання мови. З точки зору частотності аналіз Джонсон і Енслін свідчить про деякий перекид у бік чоловіків, при чому його-мова більш ніж утричі перевищує її-мову. Якісне дослідження контекстів показало, що чоловіки використовують мову, як правило, з позитивною оцінкою, тобто говорячи про щось естетично приємне, пов'язане з «простою розмовою» або з порушенням правил табу. Використання мови жінками сприймалося як позитивне рідше, і найчастіше це відбувалося в контексті виходу за рамки типових жіночих тем та тактик комунікації. Автори прийшли до висновку, що жіноча мова може інтерпретуватися як позитивна, якщо вона не є «стереотипно» жіночою. Джонсон та Енслін провели вузько сфокусоване дослідження, як з огляду на тип корпусу (лише газети), так і на критерії пошуків (його/її-мову).

Чартіс-Блек та Сіле (2009) розглянули корпус з більше 1000 інтерв'ю з людьми, які зазнали певних проблем зі здоров'ям (корпус налічував приблизно два мільйони слововживань). Орієнтуючись на гендерні стратегії, які люди застосовують, говорячи про здоров'я, автори використали онлайнвий семантичний теггер під назвою Wmatrix, який приписує семантичні коди кожному слову в корпусі і групує їх у відповідні семантичні групи. Потім дослідники застосували техніку ключових слів, щоб визначити, які концепти при відносній частотності використовуються з найбільшою різницею у порівнянні мов чоловіків і жінок. Було встановлено, що загалом чоловіки використовували такі слова, як «проблема», «важко» і «тягар» частіше, ніж жінки, посилаючись таким чином до семантичного поля «Труднощі». Якісний аналіз контексту вказав на те, що чоловіки, як правило, посилаються на хворобу як на проблему, яка потребує вирішення, а не як на досвід, який потрібно прожити. Однак, хоча автори у своїх дослідженнях розмов про хворобу і знайшли докази, що підтверджують загальноприйняті поняття маскуліності, їх аналіз також засвідчив випадки, що виступали саме проти цього.

Холмгрен (2009) провела якісний аналіз корпусу групових інтерв'ю, проведених у датському банку, з метою зосередження уваги на використанні гендерно-маркованих метафор. Хоча метафори були відібрані вручну, Холмгрен використала в якості референтних два корпуси датської мови: Корпус 90 та Корпус 2000, що складаються з більше 50 мільйонів датських слововживань. Метафоричні вирази, знайдені в її даних, потім шукалися у великих референтних корпусах для підтвердження або спростування значень, в яких вони були вжиті у корпусі групових інтерв'ю. Так, слово «udenom» («навколо, коло») було метафорично використане учасником чоловічої групи, коли йшлося про вигляд вагітної жінки. Холмгрен знайшла таку саму ШЛЯХ-метафору в референтному корпусі, що вказує на її відносну стабільність і існування за межами конкретної гендерно-ізолюваної спільноти. Проте інша метафора: *hønsøgård / høns* («курячий біг» в значенні «балаканина») була використана чоловіком-учасником, щоб описати жінок як нерозумних і постійно балакаючих. Контрольний корпус не дав жодного випадку вживання цієї метафори у такому значенні, що дозволяє припустити, що вона «виникла в дискурсивній ситуації, де респондент мав на меті приписати ролі переваги для чоловіків та неповноцінності для жінок». Робота Холмгрен (2009) відрізняється тим, що в дослідженні застосовується корпусний аналіз з метою як аналізу гендерно-маркованої мови, так і способу репрезентації гендерних уявлень. Холмгрен цікавиться всіма проявами гендерних стереотипів, виявлених за допомогою метафоричних конструкцій, які вживаються в розмові, але вона також зважає на стать комуніканта. Отже, коли мова йде про метафори, що стосуються «обходу» вагітних жінок на роботі, або про вислови про жінок як «балакучих» курчат, то примітно, що такі розмови ведуть саме чоловіки.

Бейкер (2010) використав чотири референтні корпуси письмової англійської мови з метою вивчення діахронічних змін частотності вживання не/сексистської мови протягом ХХ століття. Він знайшов докази того, що використання чоловічих займенників знизлося з 1960-х років, тоді як кількість жіночих займенників трохи зросла, хоча чоловічі займенники завжди були частішими у будь-якому періоді часу. Подібні закономірності були знайдені стосовно таких іменників, як *man* і *woman*. Знайдено свідчення того, що гендерно-інклюзивні слова, такі як *police officer*, стали використовуватися починаючи з 1990-х років, хоча такі терміни все ще були відносно рідкісними та часом вживалися з насмішкою в даних корпусу. Хоча звертання

Ms, яке позбавлене вказівки на заміжній статус жінки, не розглядається як альтернатива звертанням *Mrs/Miss*, які мають цю вказівку, Бейкер зазначає, що кількість звертань *Mr* різко знизилася з 1930-х років та припускає, що факт нерівномірності відповідних форм звертань може бути пов'язаний скоріше з відмовою від вживання *Mr*, ніж із впровадженням звертання *Ms*. Бейкер також розглянув прикметники-кологати до слів *man* та *woman*, показуючи, що хоча деякі негативні стереотипи (такі як «жінки є нестабільними») зменшуються, дані корпусу усіх часових зрізів містять посилення на впливових і сильних чоловіків, що не має відповідника для жінок, які постійно описуються з точки зору їх привабливості. Бейкер також досліджував контексти, розглядаючи різні прикметники, які вживались поряд з такими поняттями, як чоловік і жінка. Крім того, щоб об'єктивно інтерпретувати свої висновки, він спирався на інші джерела інформації, такі як статистичні дані про укладання і розрив шлюбних стосунків (коли мова йшла про способи звертання, наприклад, *Mrs*) та реакції на «політичну коректність» у 1980-х роках у Великобританії при розгляді популярності суфікса *-person* [2].

Таким чином, виникає складна картина взаємодії британського суспільства з сексистською мовою, яка свідчить на користь певної рівності, але також говорить і про те, що певні уявлення та упередження продовжують утримуватися.

Останнім дослідженням у колекції, описаній П.Бейкером, є Кінг (2011), який створив корпус комунікативних ситуацій з онлайн-чатів, взятих на веб-сайті геїв. Для того, щоб проаналізувати цей корпус, Кінг створив списки слів (wordlist – список всіх слів, вжитих у корпусі, поряд з кількістю їх вживань) для кожного чату, а потім розділив слова по категоріях на основі їхньої функції. Це привело до створення ряду основних категорій вжитого мовного матеріалу, включаючи назви таборів-угруповань, використання французької мови, заміна та реприватизація статі. На ці категорії Кінг посиляється як на аспекти існування угруповань, які виражають внутрішню солідарність та ідентичність. Кінг також стверджує, що тематичні чати (chatrooms) виникають з метафор місця і що учасники задіяні у комунікацію на основі спільних фантазій, з метою дослідження цих місць. Він демонструє це за допомогою таблиці конкордансу часто використовуваної фрази *in here*, яка показує, як чисельні учасники формують відчуття того, що вони знаходяться в тривимірному просторі під час використання чату [3].

Описані дослідження демонструють різноманітні способи використання корпусів для підтримки гендерних досліджень у мові. Чартис-Блек і Сіле (2009) та Кінг (2011) цікавляться використанням мови певними групами, а аналіз корпусу дозволяє їм визначити найважливіші та частотні патерни в межах дуже великих колекцій усних інтерв'ю та онлайн-даних відповідно. Виконане вручну, таке завдання було б теоретично можливими, але набагато більш тривалим та потенційно схильним до помилок, а також інтерпретованим відповідно до ідеологічних упереджень дослідника. Кінг використовує методи корпусних досліджень, щоб зосередити увагу тільки на одній групі, геях, і його аналіз показує, як вони спільно беруть участь у створенні самоідентичності, не звертаючись до порівняльного підходу. З іншого боку, Хартис-Блек та Сіле працюють в парадигмі порівняння чоловіків і жінок, і їх дослідження показують неоднозначну картину, де деякі висновки підтверджують попередні твердження щодо чоловічої мови, а інші суперечать їм. Таким чином, корпусний аналіз не означає матеріалізацію парадигми гендерних відмінностей, але може натомість запропонувати більш складні та різноманітні перспективи щодо взаємозв'язку між використанням мови та ідентичністю.

Малазійська дослідниця Kesumawati A. Bakar у роботі *Attitude and Identity Categorizations: A Corpus-based Study of Gender Representation* вивчає питання вербалізації гендерної ідентичності та вербального вираження ставлення до фемінності та маскулінності в онлайн-приватних оголошеннях неодружених дорослих, зібраних з ресурсу *my.match.com*. Методологія передбачає комбінацію методів корпусної лінгвістики та дискурсивного аналізу текстів. Корпус з 65659 слововживань обумовив використання автоматизованого, напів-автоматизованого та ручного аналізу патернів лінгвістичних ресурсів. Дані анотувалися за допомогою програмного забезпечення для обробки корпусів під назвою *Wordsmith* (Scott 2004). Списки частотності та конкордансів стали основою, на якій визначалися дискурсивні теми, які конструюють різні типи гендерної онлайн-ідентичності. Отримані дані показали яскраво-виражені стереотипні прояви маскулінності та фемінності, які формують домінуючі форми гендерної ідентичності. З іншого боку, були виявлені також і суперечливі дані. Два види чоловічої ідентичності, які засновані на частотності ідентифікаційних категоризацій та особливостей ставлення, є «підтягнутий та спортивний чоловік» та «чутливий молодий чоловік». Два домінуючих типи жіночої ідентичності, виявлених у даних жінок, є «ласкава жінка» і «владна жінка». Ці ідентичності формуються різними типами і поєднаннями дискурсів: це дискурс зовнішнього вигляду (тобто розмір, привабливість, підтягнутість), сфера занять (досягнень, діяльності), емоцій (щастя, афективна особистість), інтересів (спорт, спілкування) [4]. Інструменти корпусної лінгвістики дозволяють об'єднати ці дані у моделі мовної поведінки, що описують вищезгадані чоловічі та жіночі ідентичності.

Аналіз цієї підбірки наукових розвідок показує цілий спектр різноманітних методик, які може запропонувати дослідникам корпусний аналіз для проведення гендерних досліджень мови. У сукупності, згадані розвідки використовують списки частотних слів (частотності), аналіз конкордансу, списки кологатів, аналіз ключових слів, семантичне тегування, також були задіяні референтні корпуси для перевірки типовості використаних одиниць та якісного співставлення з корпусами, які складаються з більших масивів текстів.

Залишається зазначити, що не існує жодного затвердженого способу проведення корпусного аналізу – він потребує ретельного вивчення низки методів, а потім відбору найбільш підходящих з них для вирішення певних дослідницьких завдань на зібраних корпусах.

Література:

1. Порівняльні стратегії в сучасній українській художній прозі (корпуснобазований підхід) // Граматичні студії : зб. наук. праць / Донецький нац. ун-т; наук. ред. А. П. Загнітко. – Вінниця : ТОВ «Нілан-ЛТД», 2015. – С.194-201.
2. Baker, P. (2010) 'Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English.' *Gender and Language* 4.1: 125-129.
3. Baker, P. (2013) Introduction: Virtual Special Issue of *Gender and Language* on corpus approaches // *Gender and Language*.
4. Kesumawati A. Bakar (2014) *Attitude and Identity Categorizations: A Corpus-based Study of Gender Representation* / *Procedia – Social and Behavioral Sciences* 112 pp.747-756.
5. McEnergy, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.