

Отримано: 18 січня 2020 року

Прорецензовано: 23 січня 2020 року

Прийнято до друку: 29 січня 2020 року

e-mail: anzelikayanovec@gmail.com

smal.oksana85@gmail.com

DOI: 10.25264/2519-2558-2020-9(77)-139-142

Яновець А. І., Смаль О. В. Контент-аналіз політичного дискурсу: критичний огляд програмного забезпечення Linguistic Inquiry and Word Count (LIWC). *Наукові записки Національного університету «Острозька академія»: серія «Філологія»*. Острог: Вид-во НаУОА, 2020. Вип. 9(77). С. 139–142.

УДК: 81.42

**Яновець Анжеліка Іванівна,***кандидат філологічних наук, доцент кафедри української та іноземної лінгвістики,  
Луцький національний технічний університет***Смаль Оксана Віталіївна,***кандидат філологічних наук, завідувач кафедри української та іноземної лінгвістики,  
Луцький національний технічний університет*

## КОНТЕНТ-АНАЛІЗ ПОЛІТИЧНОГО ДИСКУРСУ: КРИТИЧНИЙ ОГЛЯД ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

У статті розглядається комп'ютеризована програма контент-аналізу тексту Linguistic Inquiry and Word Count (LIWC) для дослідження взаємозв'язку між політичним дискурсом та особистістю політичних діячів. Аналіз мовних зразків свідчить, що в контексті політичних мовленнєвих подій, таких як інтерв'ю в ЗМІ, політики роблять унікальний вибір лексичних одиниць, який може трактуватися як прояв певних рис особистості. Однак, незважаючи на значимість результатів, існують певні обмеження у використанні комп'ютеризованих методологій для аналізу політичного дискурсу, такі як лімітована інтерпретаційна здатність програмного забезпечення для розуміння прагматичного та контекстуального використання мови.

**Ключові слова:** LIWC; політичний дискурс; особистість; політична лінгвістика.

**Angelika I. Yanovets,***PhD, Associate Professor of Ukrainian and Foreign Linguistics Department,  
Lutsk National Technical University***Oksana V. Smal,***PhD, Head of Ukrainian and Foreign Linguistics Department,  
Lutsk National Technical University*

## POLITICAL DISCOURSE CONTENT ANALYSIS: A CRITICAL OVERVIEW OF A COMPUTERIZED TEXT ANALYSIS PROGRAM LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

The article examines and analyzes the linguistic and psychological features of political discourse using a computer-based Linguistic Inquiry and Word Count (LIWC) content analysis program to explore the relationship between political discourse and the personality of politicians. As for political discourse, it is perhaps the communicator, the linguistic personality, who plays the most important role in the communication. The linguistic personality of a politician is of particular interest in political discourse content-analysis, since it has the greatest influence on the public consciousness via mass media. Using text as a source of psychological and cognitive information has been gaining popularity. Researchers use a variety of methods to analyze texts, but Linguistic Inquiry Word Count (LIWC) has proved to be the most common technique. The analysis of linguistic patterns of political discourse shows that in the context of political speech events such as media interviews, politicians make a unique choice of lexical units, which can be interpreted as a manifestation of certain personality traits. However, despite the significance of the results, there are clear limitations to the use of computerized methodologies to make political discourse content-analysis, such as the limited interpretive capacity of software to understand pragmatic and contextual use of lexical units.

**Key words:** LIWC; political discourse; personality; political linguistics.

**Постановка проблеми.** Основна функція мови – це спілкування. Слова надають інформацію про соціальні процеси: хто має більший статус, чи добре працює група, чи хтось лукавить, чи про щирість стосунків. Вибір слова забезпечує інформацію про сприйняття певних процесів людиною. Займенники показують, як мовець ставиться до тих з ким взаємодіє. Кількість слів пояснює хто домінує в розмові та заангажованість співрозмовників. Слова і позитивно емотивні словосполучення відображають рівень згоди.

Те, як люди використовують слова у своєму повсякденному житті, може надати багато інформацію про їх переконання, страхи, модель мислення, соціальні стосунки та риси особистості. З часів робіт Фрейда про slip of the tongue до перших спроб комп'ютерного аналізу тексту дослідники почали накопичувати все більш переконливі докази того, що слова, якими ми оперуємо в мовленні, відображають наш психологічний стан.

Під час президентських виборів не можна слухати радіо, дивитись телевізор, читати газети чи дописи у соціальних мережах і не помічати, що слова політиків дуже ретельно добираються, вивчаються та постійно оцінюються. Наприклад, обмовки, невідповідності та неточності в політичному дискурсі можуть мати негативний ефект, а питання, які можуть позитивно впливати на уявлення виборців щодо того чи іншого кандидата, уникаються. Крім того, коли виборці слухають виступи та дебати, позиція кандидата щодо актуальних питань, таких як урядове регулювання, проблеми охорони здоров'я, фінансування освіти і т.д., безсумнівно, може призвести до того, що виборці проголосують за або проти кандидата. Однак, лише в останні роки вчені почали досліджувати, чи можуть конкретні мовні форми, такі як артикли та займенники, впливати на сприйняття виборцем кандидата. Щоб дослідити вплив політичного дискурсу науковці почали за допомогою комп'ютеризованих методів аналізу визначити, як мовні форми кандидатів можуть бути пов'язані з певними рисами

їх особистості. Цей аспект дослідження є важливим для розуміння того, наскільки конкретна мова політичних діячів може впливати на уявлення виборців про них, що, в свою чергу, може впливати на їх вибір у день виборів.

**Аналіз останніх досліджень.** У вітчизняному мовознавстві різні аспекти політичної комунікації частково досліджено у працях таких учених, як Ф. С. Бацивч, Т. М. Ващук, А. В. Ковалевська, Т. Ю. Ковалевська, Н. В. Кондратенко, Н. В. Кугуза, Н. В. Пеглюченко, К. С. Серажим, Л. Л. Славова, Л. М. Синельникова, О. М. Холод та ін., в той час як у зарубіжній лінгвістиці мовні питання політичного дискурсу та політичної взаємодії вивчено доволі ґрунтовно й цілісно (роботи А. М. Баранова, Е. В. Будаєва, Т. А. ван Дейка, Р. Водака, В. З. Дем'янова, М. Едельмана, О. С. Іссерс, М. В. Китайгородської, Дж. Лакоффа, Л. Лассвелла, Е. Лассан, П. Б. Паршина, Н. М. Розанової, О. А. Романова, П. Серіо, І. Ю. Черепанової, А. П. Чудінова, О. Й. Шейгал та ін.). До актуальних проблем сучасної політичної лінгвістики належать розробка та диференціація базової терміносистеми, висвітлення специфіки структурування й функціонування жанрових різновидів політичних текстів, дослідження особливостей функціонування лексичних і граматичних засобів у політичному дискурсі тощо.

Формування окремої галузі лінгвістичних досліджень, присвяченої виключно аналізу політичного дискурсу, відбулося порівняно недавно, і в ХХ ст. загальноприйнятим у сучасній науці є поява відносно нової галузі «політична лінгвістика» (А. М. Баранов, В. А. Маслова, А. П. Чудінов, О. Й. Шейгал, Н. В. Кондратенко та ін.), проте, паралельно функціонує й назва «лінгвополітологія» (Л. М. Синельникова), що вказує на міждисциплінарний характер цієї науки. На думку А. П. Чудінова, «предмет дослідження політичної лінгвістики – політична комунікація, тобто мовленнєва діяльність, орієнтована на пропаганду тих або інших ідей, емоційний вплив на громадян країни та спонування їх до політичних дій, для вироблення суспільної згоди, прийняття та обґрунтування соціально-політичних рішень в умовах множинності точок зору в суспільстві» [2, с. 6]. Ми поділяємо думку дослідника про те, що основною функцією політичної комунікації є боротьба за політичну владу, тому вона здатна емоційно впливати на адресата, трансформуючи наявну у свідомості людини політичну картину світу [2, с. 7]. Таким чином, ідеологічний потенціал та маніпулятивний вплив політичної комунікації визнано багатьма мовознавцями.

На думку Paul Chilton «Поділяючи традиції західної політичної думки, насправді існує думка, що мова та політика тісно пов'язані на фундаментальному рівні» [4]. Хоча в межах наукових дисциплін зв'язок між політикою та мовою незаперечний, вчені використовують безліч аналітичних інструментів, які базуються або на дискурсивних дослідженнях, або на певній комп'ютеризованій парадигмі (тобто LIWC), щоб вивчити політичний дискурс, особливо для встановлення зв'язку між словами політичних лідерів та їхніх особистостей.

Незважаючи на одну суттєву перепону у проведенні досліджень політичного дискурсу – обмежений прямий доступ до політиків, David Winter стверджує, що сучасні дослідники можуть вивчати політиків «на відстані» за допомогою використання якісних та кількісних методик, таких як контент-аналіз. Більше того, David Winter вважає, що конкретні методи дослідження дискурсу, такі як аналіз контенту, можуть проектувати особистісні риси політиків [10]. Наприклад, Mats Ekström виявив за допомогою контент-аналізу 19 прес-конференцій, що Джордж Буш, перебуваючи на посаді, використовував перерви та жарти, щоб тримати під контролем прес та корпус тіла, одночасно створюючи імідж впевненого та незворушного лідера [6]. Саме через переривання запитань журналістів йому вдавалося створити атмосферу самовпевненості, яку саме так трактували глядачі. Аналогічно, контент-аналіз був використаний для дослідження того, як впливає висвітлення мови політичних кандидатів засобами масової інформації на формування їх іміджу.

Інші роботи, що досліджують перетин дискурсу та особистості, включали дослідження метафоричного дискурсу, який використовували американські політики для створення публічної особистості. Зокрема, Alan Cienki використовував дискурсивний аналіз для дослідження метафоричної мови Джорджа Буша та Аль-Гора під час президентської кампанії 2000 року. Alan Cienki встановив, що хоча Буш використовував різноманітні метафори, його мова була простою на відміну від Гора, який конструював більш детальну та складну аргументацію. Alan Cienki припускає, що ці загальні мовні зразки можуть натякати на характеристики Буша та Гора як стриманих особистостей відповідно [5]. Аналогічно, через аналіз дискурсу виступів на інавгурації, Jonathan. Charteris-Black встановила, що Білл Клінтон використовував метафоричну мову для (пере) побудови своєї публічної особистості, особливо після сексуального скандалу, який вибухнув під час його другого терміну президентства. За допомогою метафор Клінтон намагався переконати у своїх добрих намірах, позиціонувавши себе як пересічного героя [3]. Так само Jonathan. Charteris-Black вивчав використання Обамою метафоричного дискурсу, виявивши, що в 19 промовах Обама використовував метафори, щоб створити образ, який є втіленням американської мрії, позиціонуючи себе як звичайну людину, яка змогла досягти надзвичайних висот, незважаючи на різноманітні труднощі [3]. Зрештою, роботи, що досліджують метафоричний дискурс політиків, свідчать, що слухачі можуть сприймати тонкі конотації цих метафор, які тим самим впливають на їх сприйняття особистостей політиків.

Хоча ці дослідження дискурсу дають захоплююче розуміння мови та особистостей сучасних американських політиків, критика методів дослідження дискурсу – відносно невелика кількість проб, що аналізуються, та потенційні інтерпретаційні ухили, які можуть виникнути під час аналізу даних – спонукали деяких учених до пошуку альтернативної парадигми для вивчення політичного дискурсу та особистості політика.

**Формулювання цілей статті, завдань.** Мета дослідження – сприяти зростаючому колу кількісних досліджень політичного дискурсу та особистостей політиків, досліджуючи можливості популярного методу комп'ютерного аналізу тексту, відомого як «Linguistic Inquiry and Word Count» (LIWC). Завдання цього дослідження – оцінити ефективність програми контент-аналізу тексту політичного дискурсу.

**Виклад основного матеріалу.** З метою забезпечення ефективного методу вивчення різних емоційних, когнітивних та структурних компонентів, присутніх у зразках словесної та писемної мови, соціальними психологами в Техаському університеті у 90-х роках було розроблено додаток для аналізу тексту в галузі психології під назвою «Linguistic Inquiry and Word Count» або LIWC (вимовляється «луке»). Друга версія LIWC2007 описується як «прозора програма аналізу тексту, яка рахує слова у психологічно значущих категоріях» [9]. Спочатку він використовувався в соціальній психології для дослідження зв'язку між вживанням слова та відновленням психічного здоров'я, пізніше – для дослідження задоволення від відносин, університетськими ступенями та рівнем тестостерону. Він також використовувався для відстеження колективних реакцій

на потрясіння, таких як напади 11 вересня. Творці LIWC з тих пір використовують програму для аналізу есе, виступів, публікацій в блогах та твітів тисяч людей у різних проєктах. Це призвело до того, що було проведено багато досліджень, з використанням LIWC у різних сферах: особистісне профілювання, виявлення обману, аналіз настроїв, контент-аналіз та виявлення спаму, тощо.

LIWC дає можливість класичного підрахунку слів та пунктуаційних типів, які часто використовуються як проксі для складності тексту та структурної композиції. Існують підрахунки функціональних слів, які використовуються для авторської стиліметрії:

LIWC2015 дає можливість включати зовнішні словники (різними мовами) та навіть словники, створені дослідниками. (Вбудований словник, однак, користувачі не можуть його редагувати).

LIWC отримав широке використання як альтернативна методологія в аналізі політичному дискурсі та особистості політика. Спираючись на припущення, що політичний дискурс може бути точним показником індивідуальних відмінностей особистості, Pennebaker, Chung, Ireland, Gonzales і Booth розробили і використали новий теоретико-методологічний підхід у вимірюванні особистості за допомогою мови. LIWC та інші «Стратегії підрахунку слів ґрунтуються на припущенні, що слова, якими користуються люди, передають психологічну інформацію понад їх буквальне значення і не залежать від їх семантичного контексту» [9]. Теоретичне припущення методології LIWC – це поняття про те, що частини мови можна класифікувати та кількісно оцінити, що дає можливість зробити висновки про характеристики особистості людей. У цьому методі мовні зразки обробляються шляхом узгодження окремих основ (stem) слова в межах зразків із словосполученнями у внутрішньому словнику програмного забезпечення. Після того, як кожне слово було належним чином розпізнане, воно ставиться у відповідну категорію. Наприклад, слово «active» було б ідентифіковане програмним забезпеченням і класифіковане як дієслово незалежно від суфіксації (наприклад, третя особа -s, маркер минулого часу -ed). Однією з помітних переваг використання методології підрахунку слів є те, що велика кількість мовних зразків може бути зібрана та проаналізована швидше, ніж за допомогою в інших методик, таких як аналіз дискурсу та аналіз контенту.

Навчальне дослідження, що використовує методика LIWC, проведене Pennebaker and King, яке вивчало зразки письма від широкого кола осіб, включаючи студентів коледжу, стаціонарів, які беруть участь у програмах наркоманії, та дослідників. Завдяки комп'ютерному аналізу тексту письмових зразків дослідники стверджують, що мова була надійною конструкцією для вимірювання особистості. Тобто, мова може виявляти аспекти особистості. Більше того, конкретні характеристики особистості – відтепер їх називають вимірами особистості – були створені та перевірені шляхом аналізу надійності та валідності. Один із прикладів створеного виміру особистості включає когнітивну складність або здатність формувати відмінності в мові за допомогою слів, що виражають причину та наслідки (e.g. «affect» та «depending») та обізнаність (e.g. «knowledge» та «ponder») [7]

Для перевірки комп'ютеризованого методу LIWC було проведено лінгвістичний аналіз виступів кандидатів у президенти та віце-президенти Обама та Маккейна та Байдена та Пейліна, щоб визначити, чи існують суттєві відмінності в їхньому дискурсі. Крім того, було проведено аналіз, щоб визначити, чи можна виявити мовні відмінності відповідно до позиції політичної партії та кандидата у президенти, чи віце-президенти.

Також мовленнєві зразки були проаналізовані за допомогою програми LIWC для визначення прояву таких вимірів особистості: когнітивної складності, депресія, президентство, чесність та фемінітивність.

1. Когнітивна складність – вимір когнітивної складності складається з більшої кількості попереднього показника (e.g. «appears», «depending», «mainly»), невідповідності («should», «would», «want»), виключення (e.g. «except», «unless», «without»), та заперечних слів (e.g. «neither», «never», «none») ще менша кількість слів включення (e.g. «come», «including», «plus»).

2. Депресія – вимір депресії складається з більшої кількості сингулярних посилань від першої особи (e.g. «I», «mine»), слів негативних емоцій (e.g. «mad», «struggle», «tense»), та фізичних слів (e.g. «nutrition», «physical», «sensation»), і ще менша кількість слів, що описують позитивні емоції (e.g. «cherish», «pleasurable», «thankful»).

3. Президентство – вимір президентства характеризується більшою кількістю слів позитивних емоцій, великими словами (тобто словами, що містять більше 6 букв), артиклями (e.g. «a», «an», «the») та прийменниками (e.g. «above», «on», «through»).

3. Чесність – розмір чесності складається з більшої кількості посилань на себе (e.g. «I» and «my») а також посилань на інших (e.g. «he», «oneself», «she») та більшої кількості слів виключення. Однак чесні люди, як правило, використовують менше рухомих слів (e.g. «act», «drive», «enter») та негативні емотивні слова.

5. Фемінітивність – вимір фемінітивності особистості індивіда складається з більшої кількості посилань на інших, а також позитивних емоційних слів, ще меншої кількості прийменників, статей, великих слів, заперечень, ненормативної лексики (e.g. «ass», «hell», «suck»), і обмежена згадка про гроші (e.g. «bill», «cost», «money») і слова, пов'язані з числом (e.g. «billion», «hundred», «seventy»).

Програми аналізу тексту за своєю суттю є філософськими. Тобто, все програмне забезпечення лежить в основі теорій та припущень, а у випадку програмного забезпечення для аналізу тексту – це теорії мови та припущення про те, що має сенс. Важливо, щоб перед його використанням ми визначили та критично оцінили теоретичні основи програми аналізу тексту. Нижче наведено деякі основні особливості LIWC та теоретичні припущення сприйняття цієї методики.

LIWC рахує слова. Виходячи з контексту, в якому створено LIWC, основне припущення полягає в тому, що частота слова може сказати нам щось про людину або про зміст або тон тексту. Вторинне припущення полягає в тому, що комп'ютерна програма ідеально підходить для виконання цього завдання. LIWC розглядає лише окремі слова. Роблячи це, LIWC припускає, що слова мають значення ізольовано. Існує також неявне припущення, що неточності внаслідок заперечення, порядку слів, часток (наприклад, у фразових дієсловах), неоднозначності словосполучень, типу дискурсу та інших факторів, що залежать від контексту, є незначними або несуттєвими

Програма співставляє слова з наявними у словниках. Створюючи словники апіорі, а також знаходячи і рахуючи лише ті слова, які відповідають їм, будуть ідентифіковані цікаві або релевантні частини тексту. Оцінка, що відповідає мітці словни-

ка, також буде відповідати передбачуваному семантичному значенню цього словника. Однак, семантичне значення є більш важливим, ніж частота слів.

**Висновки і перспективи подальших розвідок.** Метою цього дослідження було оцінити загальнозживану програму аналізу тексту в політичному дискурсі. Незважаючи на результати цього дослідження, що виявляють змістовні відмінності в мові кандидатів, а також їх політичних партій, це дослідження показало, що LIWC, хоча і ефективно, аналізує мовні зразки майже миттєво і менш суб'єктивне, ніж дискурсивні дослідження, обґрунтованість отриманих результатів цієї парадигми сумнівні. Навряд чи дослідження, що використовують цю методологію, точно вимірюють взаємодію між дискурсом та особистістю, оскільки ці методи обмежені кількісними показниками, а також неможливістю розглянути контекстне використання мови – ключовий фактор, який мовець і адресат зможуть інтерпретувати як свій вибір.

У дослідженні було охарактеризовано LIWC, щоб продемонструвати деякі основні переваги та потенційні помилки програмного забезпечення для аналізу тексту. Зважаючи на зростаючу популярність комп'ютеризованого аналізу тексту, важливо, щоб програми для підрахунку слів, такі як LIWC, використовувались з обережністю, особливо дослідниками не лінгвістами та за межами процесу спілкування. Користувачам слід розуміти, що все, що є «об'єктивним» та філософсько-нейтральним, все програмне забезпечення засноване на теоретичних припущеннях, деякі з яких є предметом постійних дискусій. Щодо ефективності програми LIWC, що базується на підрахунку слів, як, виявляється, що цей підхід має ряд обмежень для аналізу контенту. Однак, якщо і розробник програми, і кінцевий користувач уважні та рефлексивні у розгляді теоретичних припущень, такі обмеження можуть бути усунені. Вважаємо, що LIWC працює краще в поєднанні з іншими, більш якісними інструментами аналізу, а контекст є першорядним при вимірюванні значень у текстах.

Перспективи подальших розвідок бачаєм у дослідженні використання LIWC для ідентифікації, аналізу та класифікувати ідіоматичних виразів з основоположними прагматичними функціями у політичному дискурсі.

#### Література:

1. Семенюк О. А., Парашук В. Ю. Основи теорії мовної комунікації: навч. посіб. К.: ВЦ «Академія», 2010. 240 с.
2. Чудинов А. П. Политическая лингвистика: [учеб. пособие]. М.: Флинта : Наука, 2008. 256 с.
3. Charteris-Black, Jonathan. 2011. *Politicians and Rhetoric: The Persuasive Power of Metaphor*. Basingstoke: Palgrave MacMillan.
4. Chilton, Paul. 2004. *Analyzing Political Discourse: Theory and Practice*. London: Routledge.
5. Cienki, Alan. 2004. "Bush's and Gore's Language and Gestures in the 2000 US Presidential Debates: A Test Case for Two Models of Metaphors." *Journal of Language and Politics* 3 (3): 409–440. doi:10.1075/jlp.3.3.04cie
6. Ekström, Mats. 2009. "Power and Affiliation in Presidential Press Conferences." *Journal of Language and Politics* 8 (3): 386–415. doi:10.1075/jlp.8.3.03eks.
7. Pennebaker, James W., and Laura A. King. 1999. "Linguistic Styles: Language Use as an Individual Difference." *Journal of Personality and Social Psychology* 77: 1296–1312.
8. Pennebaker, James W., Matthias R. Mehl, and Kate Niederhoffer. 2003. "Psychological Aspects of Natural Language Use: Our Words, Our Selves." *Annual Review of Psychology* 54: 547–577. doi:10.1146/annurev.psych.54.101601.145041.
9. Pennebaker, James W., Richard B. Slatcher, and Cindy K. Chung. 2005. "Linguistic Markers of Psychological State through Media Interviews: John Kerry and John Edwards in 2004, Al Gore in 2000." *Analyses of Social Issues and Public Policy* 5 (1): 197–204.
10. Winter, David G. 2005. "Things I've Learned about Personality from Studying Political Leaders at a Distance." *Journal of Personality* 73 (3): 557–584.
11. Secretary of State Hillary Clinton's testimony on Benghazi, 2012. URL: <http://www.foxnews.com/politics/2013/01/23/transcript-secretary-state-hillary-clinton-testimony-before-senate-committee/>
12. Barack Obama. Address at the Transfer of Remains Ceremony for the Victims of the Attacks on the U.S. Consulate at Benghazi, 2012. URL: <http://www.americanrhetoric.com/speeches/barackobama/barackobamabenghazivictims.htm>