

ФОРМАЛЬНИЙ АНАЛІЗ ПОЛІТИЧНОГО ДИСКУРСУ

У дослідженні здійснено кореляційний діахронічний аналіз дискурсивної структури політичного процесу України. Джерелами дослідження визначені тексти Послань Президента України до Верховної Ради за 2009-2020 роки. Показники були розраховані за допомогою мови програмування R, із застосуванням відкритих аналітичних пакетів, виходячи із розробленого автором алгоритму. Наведено засадничі положення щодо аналізованого корпусу текстів як такого, що може вважатися репрезентативним з огляду на цілі дослідження. Зафіксовано рівень кореляції, який із застосуванням описуваних перед підготовчих етапів коливається в окремих випадках від 0,45 до 0,82. Визначено необхідний мінімальний рівень 0,4 як такий показник кореляції, що може розглядатися в якості значного.

Визначено ключові дискурсивні патерни (ДП) виражені в наборі понять та категорій, які використовуються для опису окремих тем в рамках текстів. Побудовано таблицю ключових елементів кожного дискурсивного патерну із огляду на їх статистично-ймовірнісне використання в аналізованому корпусі. Побудовано таблицю ієрархізованої вживаності ДП через діахронічний аналіз використання в текстах. Визначено, що для більшості аналізованого масиву присутнє співпадіння на першому рівні ДП (із виключенням в тексті 2013 року). Зазначено, що формальні результати проведеного дослідження можуть вважатися доведенням тези про безсуб'єктність політичного процесу, який фактично є самоактуалізацією дискурсу. Окреслено подальший розвиток автоматичного формального дискурсивного аналізу, що полягає в уточненні та розкладанні елементів ДП через окремі тексти корпусу із визначенням співпадінь та розходжень в семантичному наповненні елементів та побудову однорангової мережевої моделі синхронічної дискурсивної структури для кожного із визначаємих етапів.

Ключові слова: дискурс, формальний дискурс-аналіз, дискурсивний патерн, політичний процес, латентно-семантичний аналіз.

Застосування дискурсивного аналізу на чинному етапі є вкрай важливим академічним завданням. Адже, з одного боку, дискурсивний аналіз досить розповсюджений метод в сучасній науці, з іншого, його використання здебільшого носить супутній характер, що дозволяє лише частково розкрити сутність методу та відносну новизну тих даних, які можна отримати при суворому його використанні. Окрім того, також важливим є саме формальний аспект такого аналізу, позитивність застосування якого вбачається нам не тільки в самій структурі отримуваних даних, а також у тому, що ступінь їх верифікуємості дозволяє використання такого роду аналізу в різного роду дослідженнях, як в якості часткового методу, так і загальної методології.

В рамках попередніх робіт ми розглядали як теоретичні, так і практичні аспекти та можливості формального аналізу політичного дискурсу. Метою даного дослідження є часткова реалізація розроблених алгоритмів автоматичного дослідження дискурсивної структури на матеріалах політичних текстів. Нами були взяті найбільш виразні тексти, а саме послання Президента до Верховної ради України, повторюючись зазначимо, що вони досить об'ємні для того, щоб надати необхідний мінімальний обсяг текстових одиниць для аналізу, а також досить виразні в плані опису як більш седиментованих шарів, так і виразні в плані демонстрації політичного порядку денного щодо тем, існуючих в суспільстві. Для представлення діахронічного аналізу нами були обрані Послання Президента України до Верховної Ради за наступні роки: 2009, 2011, 2013, 2015-2020 роки. Так, говорячи про хронологію аналізованих послань / звернень, ми відбирали тексти виходячи з наступних положень: – представити структуру дискурсу в межах таких часових інтервалів, які б відображали внутрішньополітичні зміни, а саме при аналізі використовуються тексти чотирьох Президентів, а крім того, очевидними і таким, що не вимагають пояснень нам

уявляються зміни як територіальні, так і інформаційні, що відбулися в політичному просторі й до певної міри, як нам здається, з неодмінністю впливають на політичний процес, а отже, й на дискурсивну структуру; – відобразити ключові віхи в потенційно можливих точках зламу дискурсивної структури, які пов'язані з фактами, описаними вище.

Сам дискурс виражений в слабо ієрархізованій структурі понять (категорій), які можна вивести практично з будь-якого політичного тексту (звісно, беручи до уваги його об'ємну характеристику), відносно незалежно від його жанрових, тематичних та ін. особливостей. Домінуюча лінія опису і дій в рамках політичного процесу іменується «гегемонний» дискурс.

Ми припускаємо, що дану «структуру» дискурсу можливо виокремити і розібрати за допомогою формально-лінгвістичних інструментів аналізу, представивши її у вигляді однорангової мережі, в якій кожен елемент є як організовуючим, так і організованим іншими елементами, перебуваючи з ними в складній структурі відносини спів підпорядкування. Сама мережа, таким чином, в максимально можливій надлишковій формі описує структуру політичного дискурсу всередині окремого символічного об'єднання. Даний підхід, дозволяє нам розглядати політичний процес не з точки зору дій окремих політичних акторів, що керуються приватними цілями і завданнями, а як процес, підпорядкований логіці дискурсу, мета утворення, що визначає логіку дій еліти.

Окрім того, слід сказати про існуючий парадокс, який виражається в тому, що обсяг аналізованих текстових даних (а саме послань Президента України до Верховної Ради України) є і вичерпним і недостатнім одночасно. З одного боку, подібного роду тексти можуть бути дещо недостатніми для абсолютно вичерпного уявлення про дискурсивну структуру, залишаючи деякі «осередки» незаповненими, з іншого боку текстуальний обсяг достатній для того, щоб вважати показник «чистої» (під цим ми розуміємо такий вид взаємозв'язку документів, який представлений без тих мовленнєвих норм, які затверджені і продиктовані використовуваною мовою) кореляції понад 0,4 як значний, а понад 0,5 як дуже високий.

Таким чином, використовуючи розроблений алгоритм, здатний аналізувати структуру дискурсу в окремо взяті проміжки часу, ми маємо можливість провести діахронічний аналіз за допомогою складання графіків кореляції текстів на основі латентного розміщення Діріхле. Даний метод, розроблений по більшій мірі у зв'язку із зростанням необхідності поліпшення роботи інформаційного пошуку і машинного навчання, втім, з деякими застереженнями, його можна використовувати і в контексті наявних завдань.

Матеріали та метод

Сутність методу, покладеного в основу проведеного дослідження полягає в поєднанні постструктуралістської теорії дискурсу з методом математичного розкладання, заснованим латентно-семантичному аналізі і латентному імовірнісному розподілі Діріхле.

Останній, вперше метод був представлений як Графова модель, Девідом Блеєм, Ендрю Ином і Майклом Джорданом в 2003 році, як такий, що дозволяє тематизувати тексти з метою виявлення ключових слів по відношенню до кожної з тем [2]. Відносно нашого дослідження, відзначимо, що його використання дозволяє з деяким рівнем допущення складати тематичне відображення дискурсивних патернів для окремих текстів в корпусі аналізованих документів. Таке тематичне моделювання (далі формування дискурсивних патернів) може бути використано для виокремлення найбільш ключових тем у межах текстів з подальшим розкладанням на окремі знакові елементи (опорні точки), які, за допомогою використовуваного алгоритму, а саме завдяки засобам дистрибутивної семантики і латентно-семантичного аналізу, можливо проаналізувати на порівнянність і включеність інших термів у формування ключового знаку (яку, як нам здається, можна вважати трансформованою формою ланцюгів еквівалентності) [3].

Використовуючи розроблений нами алгоритм, а також певні формули і моделі, ми можемо провести наступне діахронічне дослідження з можливістю порівняння в текстуальній складовій елементів, використовуваних в дискурсивних структурах різних років. Це не означає, що ми припускаємо повну зміну структури дискурсу щороку (що підтверджується тим, що з 2009 по 2015 роки ми брали тексти послань кожні два роки, припускаючи, що структура залишиться незмінною в силу особливостей політичного процесу), проте, ми не припускаємо також, що дискурсивна структура змінюється зі зміною президентів, навіть більше, кажучи про безсуб'єктність в межах політичного процесу, що ми уточнимо надалі, скоріше навпаки, зі зміною політичних суб'єктів не посту лідера, дискурс не має тенденції змінюватися відразу, скоріше, для цього необхідні деякі як зовнішньополітичні, так і внутрішньополітичні трансформації, що «змушують» гегемонний дискурс, інкорпорувати нові елементи, зберігаючи і видозмінюючи структуру, з метою отримання можливості

опису існуючого політичного порядку. Прикладом відсутності кардинальних змін у формальній структурі дискурсу служить відносно високий ступінь кореляції звернення в 2020 році з текстами попередніх послань, але про це йтиметься далі.

Слід сказати, що дані, які ми будемо наводити у наступних таблицях мають повністю відфільтровані показники за принципом застосування стоп-словника (видалення всіх слів, які не несуть семантичного навантаження і служать скоріше засобами синтаксису, об'єднання окремих елементів мови у зв'язний текст), крім того, матеріали текстів, при використанні латентного розподілу Діріхле, не були відфільтровані за допомогою tf-idf механізмів (про які ми говорили у попередніх публікаціях, тут зазначимо, що мова йде про зворотну кореляцію значущості термів по відношенню до загальної вживаності в значній частині корпусних документів), що дозволяють більш детально відслідковувати елементи, які є постійно повторюваними, але не включеними в список стоп-словника, оскільки набір проаналізованих текстів не був великий (а саме – 8 текстів), а, крім того, стосується умовно чітко заданої теми, тому даний коефіцієнт лише знижував рівень достовірності результатів кореляції (але підкреслимо його необхідність на більшому масиві текстів, які можуть бути різноспрямовані в плані стилю, жанру (в розумінні лінгвістичного аналізу) і т. і.). Отже, маємо зазначити, що набір використовуваних перед підготовчих технік дещо знижує рівень кореляції окремих документів, оскільки прибирає всі слова, які не відносяться до патерного розподілу, при цьому, показник кореляції тим вище, чим ближче він до одиниці.

У табл.1 наводяться значення коефіцієнтів кореляції, що характеризують взаємну наближеність текстів «звернень» Президента. Аналізуючи її, ми можемо відзначити деякі значущі позиції щодо дискурсивних змін у межах заданого часового проміжку, а саме 2009-2020 років.

Таблиця 1

Матриця коефіцієнтів кореляції текстів «звернень» президентів України

	2009	2011	2013	2015	2016	2017	2018
2011	0.65770						
2013	0.51909	0.45373					
2015	0.70495	0.64261	0.46993				
2016	0.68861	0.61490	0.43150	0.79331			
2017	0.72029	0.65285	0.53452	0.82618	0.81640		
2018	0.69219	0.61984	0.49230	0.78319	0.80788	0.81537	
2020	0.67759	0.65577	0.48693	0.72410	0.69007	0.72557	0.73480

Виходячи з табл.1, суттєво значущим ступенем кореляції, що не дивно, відрізняються роки, які на посаді Президента проводила одна людина, а саме часові проміжки 2011-2013, 2015-2018. Примітно так само, що рівень кореляції високий і щодо президентської каденції 2015-2018 року і часового інтервалу 2019-2020 років. Тут, для чистоти дослідницьких даних, нам би хотілося відзначити, що так чи інакше подібні тексти з необхідністю будуть корелювати здебільшого хоча б щодо структури, але в той же час, показники на фіксованому рівні все ж залишаються дуже високими, а тому можуть бути сприйняті як показові. Виходячи з такого рівня кореляції можемо зробити щонайменше два ключових висновки: перший полягає в тому, що високий рівень кореляції дозволяє говорити про наявність відносно незмінного кола питань, що функціонують в рамках політичного поля в різних часових проміжках; другий формується виходячи з епістемологічних засад застосованої дискурсивної теорії, і полягає в тому, що політичний суб'єкт не може бути сформульований як такий, що дискурсивну структуру продукує. Це, з одного боку може бути виведено з даних щодо кореляції, наприклад, текстів 2015-2018 років із текстами 2019-2020 років. Адже, їх співпадіння досить дивний момент, оскільки обрання Президента в 2019 році було позначено «абсолютно новим» кандидатом, який не включений в існуючий політичний устрій і позиціонувався як людина, здатна реорганізувати ту дискурсивну модель, яка стала невідривно пов'язуватись із чинними політичними акторами в рамках символічного простору політичного процесу. З іншого боку, це більшою мірою підтверджує нашу гіпотезу, що була взята як певна епістемологічна основа, щодо наявності гегемонної дискурсивної структури. Адже, включеність в гегемонно організований дискурс означає, що окремо взятий політичний актор, що покладається у своїй діяльності на ряд стійких ментальних диспозицій, в кінцевому рахунку існує в рамках фрагментарної ідентичності,

обмеженої дискурсом. Суть, будь-які дії Президента, отже, можуть бути розглянуті і, навіть більше, прогнозовані за допомогою виокремлення структури дискурсу. Це побічно підтверджує наявність певної форми того, що ми можемо назвати «габітусним заломленням», перед задаючим позиції, з яких може виступати той чи інший політичний актор. Таким чином, це побічно підтверджує гіпотезу вживаної форми дискурсивного аналізу щодо того, що політичний дискурс носить безсуб'єктний характер, адже при такій високій схожості дискурсивних структур, які, крім іншого залежать не стільки від конкретних політичних акторів, а скоріше від незвичайного синтезу артикуляційних актів, що трансформують дискурс і від зовнішніх обставин, що вимагають акту дискурсивного визначення, рівень «усвідомленого» впливу окремих політиків на дискурсивну структуру може бути визначений як незначний.

Проводячи подібного роду аналіз ми маємо можливість виділити дискурсивні патерни в межах як окремого тексту, так і групи текстів, що відносяться до певного корпусу з тим, щоб виокремити основні використовувані патерни в текстах (а, отже, з певним ступенем допущення і екстраполяції методом індуктивного аналізу) і в межах дискурсивної структури. ДП, в свою чергу, складаються з набору ключових елементів ланцюга. Спираючись на доведеність положення, згідно якому в на кожному рівні смислового ланцюга значущими елементами можуть бути близько 7 елементів [1, с. 204; 4, с. 90-92; 5, с. 162], ми виходили з того, що має бути близько 7 ключових елементів "верхнього рівня". Теоретично на кожному смислового рівні кожне поняття (елемент) дезагригується на сім плюс-мінус три поняття (елемента), які співвідносяться зі структурою термів, що їх визначають та становлять основу для побудови дискурсивної структури щодо стійкого символічного набору.

Таблиця 2

Елементи верхнього рівня, що складають дискурсивні патерни

1	Українець, Президент, Крим, Армія, Корупція, Медицина, Програма, Партія
2	Росія, Війна, Донбас, Сила, Територія, Боротьба, Агресія, Мова, Євросоюз
3	ООН, Революція, Право, Дискусія, Ефективність, Банк, Освіта
4	Позиція, Підтримувати, Нація, Культура, Ресурс, Соціальний, Самоврядування, Влада, Криза, Вибачити, Разом
5	Україна, Країна, Держава, Система, Світ, Влада, Реформа, Економіка, Рішення, Народ, Національний, Нова (прим. Держава)
6	Експорт, Відновлення, Оборона, Захід(ний), Кремль, Схід, Очікування, Припинення, Атака, Зсув, Більшість
7	Розвиток, Реформа, Ресурс, Виробництво, Модернізація, Інновація, Реалізація, Потенціал, Можливість
8	Рух, Сценарій, Інформаційний, Перемога/перемагати, Гарантувати, Правий, Свобода, Втручатися, Поліпшення, Досвід
9	Президент, (Верховна) Рада, Закон, Суд, Парламент, Конституція, Порядок
10	Правило, Зв'язок, Газ, Субсидія, Тариф, Монополія, Гарантія, Децентралізація, Дім, Приватизація

Таким чином, ми отримуємо таблицю з ключовими дискурсивними патернами, які можна представити як вузлові точки в межах однорангової системи. Саме за допомогою моделі однорангової системи ми намагаємося вирішити питання багатоступовості структури. З одного боку, ми залишаємо на рівні теоретичного допущення деякі елементи структури, які більшою мірою седиментовані в політичному дискурсі і тому, в меншій мірі піддаються трансформації і реорганізації, будучи плаваючими знаками, які організовують гегемонний дискурс. При цьому, боротьба за їх зміну і наповнення новими ланцюгами еквівалентності може відбуватися на маргінальних ділянках дискурсивного простору і не підлягає перегляду на більш високих рівнях (наприклад. політичного порядку). Отже, такі знаки, в цілому, поділяються переважною більшістю політичних сил і окремих акторів.

Наступним кроком дослідження є побудова таблиці, що демонструє розподілення дискурсивних патернів в рамках окремих текстів звернень Президента (табл.3).

**Розподіл дискурсивних патернів у зверненнях президентів України
за рівнем пріоритету**

<i>Рівень пріоритету</i>	<i>2009</i>	<i>2011</i>	<i>2013</i>	<i>2015</i>	<i>2016</i>	<i>2017</i>	<i>2018</i>	<i>2020</i>
1	5	5	9	5	5	5	5	5
2	4	7	5	2	2	2	2	1
3	9	4	4	10	6	3	8	2
4	7	3	7	6	7	9	9	7
5	10	2	2	4	9	7	6	9
6	3	9	8	9	4	8	1	3
7	1	1	6	1	1	6	7	10
8	8	6	3	3	8	4	3	4

Дана таблиця, може бути спочатку трактована в тому, що на рівні первинного політичного порядку дискурсивна структура практично у всіх часових інтервалах стабільна, і має лише деякі допустимі коливання, в межах «контрольованого» відхилення (наприклад., 2013 рік). Але, виходячи з використовуваної в якості методологічної бази дискурсивної теорії пост структуралізму найбільш важливим є не тільки відповідність ключових позицій, а й їх визначення через ланцюги еквівалентності. Тобто, для кожного ключового знаку, використаного в дискурсивних патернах, ми з необхідністю повинні отримати структуру категорій, які знаходяться в найбільш високому ступеню зв'язку з окремим терміном, що дозволить більш детально зобразити як синхронічну дискурсивну структуру, так і простежити діахронічні аспекти.

Висновки

Таким чином, проведене дослідження дає засади говорити про те, що автоматичний формальний аналіз дискурсивної структури можливий та дозволяє отримувати новий формат даних для аналізу з точки зору політологічного дослідження. Також, можемо зазначити, що подальший вектор такого дослідження полягає у деталізації та всебічному розгляді окремих дискурсивних патернів, які із застосуванням наукової редукції можна визначити як вузлові точки для однорангової мережевої моделі дискурсивної структури. Він полягає у формуванні шляхом трансформованого латентно-семантичного аналізу дезагрігованих елементів для зображення більш детальної структури дискурсу. При цьому маємо зазначити, що в подальших аналітичних дослідженнях, важливо проаналізувати зміну семантичної наповненості ключових знаків, через ланцюги еквівалентності в рамках відносно зовні стабільної та повторюваної структури дискурсу.

Бібліографічний список:

1. Польовий М. А. Політичні процеси: теорія та практика моделювання : монографія. Одеса: Фенікс, 2011. 288 с.
2. Blei, D.M. Latent Dirichlet Allocation Journal of Machine Learning Research. 2003. Vol. 3. P. 993-1022.
3. Laclau E., Mouffe Ch. Hegemony and Socialist Strategy: Towards a Radical Democratic Politics / London : Verso. 1985. – 199 p.
4. Miller G. A. . The Magical Number Seven, Plus or Minus Two / 1956. Vol. 63. P. 81–97.
URL: <http://psychclassics.yorku.ca/Miller/> (дата звернення 23.01.2021).
5. Reznikova Zh. Animal Intelligence: From Individual to Social Cognition / Cambridge University Press, 2007. – 488 p.

References:

1. Polovyi M. A. Politychni protsesy: teoriia ta praktyka modeliuvannia : monohrafiia. Odesa: Feniks, 2011. 288 s.
2. Blei, D.M. Latent Dirichlet Allocation Journal of Machine Learning Research. 2003. Vol. 3. P. 993-1022.
3. Laclau E., Mouffe Ch. Hegemony and Socialist Strategy: Towards a Radical Democratic Politics / London : Verso. 1985. – 199 p.
4. Miller G. A. . The Magical Number Seven, Plus or Minus Two / 1956. Vol. 63. P. 81–97.
URL: <http://psychclassics.yorku.ca/Miller/> (дата звернення 23.01.2021).
5. Reznikova Zh. Animal Intelligence: From Individual to Social Cognition / Cambridge University Press, 2007. – 488 p.

Kotov E. V. Formal Analysis of Political Discourse

As part of the study, a correlation diachronic analysis of the discursive structure of the political process in Ukraine was undertaken. The sources of the study were the texts of the Messages of the President of Ukraine to the Verkhovna Rada for 2009-2020. The indicators were calculated using the programming environment R, using the open analytical packages, based on the developed algorithm. There were given the basic provisions concerning the analyzed corpus of texts as ones which can be considered representative considering the purposes of the research. The level of correlation was fixed, which varied in some cases from 0.45 to 0.82 with the application of the described preparatory stage. The required minimum level of 0.4 was determined as that correlation indicator which can be considered as a significant one.

The key discursive patterns (DP) were determined in a set of concepts and categories that were used to describe individual topics within the texts. There was constructed the table of key elements of each discursive pattern considering their statistical-probabilistic use in the analyzed corpus. The table of hierarchical usability of DP through the diachronic analysis of use in texts was constructed. It was determined that for the majority of the analyzed array there was a coincidence at the first level of the DP (excepting the text of 2013). It was indicated that the formal results of the study can be considered as a proof of the thesis of the subjectlessness of the political process, which was in fact a self-actualization of discourse. There was described the further development of automatic formal discourse analysis, which meant the clarifying and decomposing the elements of the DP through separate texts of the body with the definition of coincidences and differences in the semantic content of elements and building a peer-to-peer network model of synchronous discursive structure for each stage.

Keywords: *discourse, formal discourse analysis, discursive pattern, political process, latent-semantic analysis*