

ВИЗНАЧЕННЯ ПАРТІЙНИХ ПОЛІТИЧНИХ ОРІЄНТАЦІЙ НА ОСНОВІ НЕСТРУКТУРОВАНИХ ТЕКСТОВИХ ДАНИХ: МЕТОДОЛОГІЧНА ОЦІНКА ПІДХОДІВ

У статті представлено чотири найпопулярніші на сьогодні підходи до визначення орієнтацій щодо обраної проблеми, які передбачають аналіз висловлювань у формі неструктурованих текстів: експертні оцінки, якісний контент-аналіз, кількісний словниковий аналіз та шкалювання без застосування кодувальної схеми. Проведено порівняння вказаних підходів на предмет валідності та надійності продукованих результатів, технічної зручності застосування, ефективності та ресурсомності.

Ключові слова: контент-аналіз, політика, міжнародна політика, зовнішньополітичні орієнтації, машинна обробка мови, українська політика, партії, програми, експертні оцінки, Україна, Росія.

Дослідження політичної конкуренції в сучасних демократичних країнах часто пов'язані з визначенням позицій політичних акторів щодо одного чи більше вимірів стратегії розвитку суспільства. Таким чином, виявлення валідних та надійних індикаторів політичних позицій є однією з центральних проблем у дослідженнях політичних (владних) еліт та електоральної поведінки. Тому не дивно, що існує велика кількість наукової літератури, пов'язаної з методологічними підходами до визначення політичних позицій. Один із порівняно недавніх і досить повних оглядів сучасних методів у цій сфері опубліковано в журналі «Electoral Studies» в 2007 році [18, с. 23–38]. Огляд містить оцінку переваг та недоліків експертного опитування щодо позицій партій (апріорного, без аналізу текстів програм), контент-аналізу передвиборчих програм і кількісних опитувань громадської думки.

На пострадянському просторі і в Україні зокрема є суттєві прогалини саме в методологічному аналізі досліджень політичних позицій. Найбільшу увагу зосереджено на дослідженні установок, орієнтацій, громадської думки масоподібних суб'єктів (електоральних груп) та індивідів. В цьому напрямі є достатня кількість як емпіричних досліджень, так і методологічних розробок власне українського походження.

Натомість дослідження позицій політичних партій обмежуються експертними оцінками (в переважній більшості, див. наприклад [1, с. 16–19]) та якісно-кількісним контент-аналізом програм і виступів політиків, який передбачає ручне кодування вихідних даних відповідно до попередньо розроблених категорій, після чого із закодованими і «оцифрованими» даними проводять різного роду статистичні операції [3,

с. 5–14]. В Україні (як і в інших пострадянських державах) не здійснюють ані багаторічні лонгitudні дослідження політичних позицій партій за єдиною кодувальною схемою, ані словникові контент-аналітичні дослідження з автоматизованим первинним кодуванням, не використовують новітні методики шкалювання текстових документів, що й казати про більш складні й ресурсомісткі дослідження відповідності декларованих політичних позицій реальній поведінці політиків (прийняття рішень), т. зв. *roll call* дослідження [11, с. 381–396]. Видається, що такі дослідження просто не вигідні сучасним українським політичним елітам, які є основними замовниками електоральних досліджень.

Через згадані прогалини виникає дисбаланс і в теоретико-методологічному розумінні поняття «політичні позиції». У випадку кількісних досліджень установок, орієнтацій, цінностей тощо електорату та глибинних опитувань окремих індивідів є сенс розглядати «політичну позицію» з погляду, наприклад, диспозиційної концепції особистості [5, с. 47–62], як елемент взаємопов'язаної структури, що має ієрархічний характер, на вершині якої стоять ціннісні орієнтації. З іншого боку, якщо «політичну позицію» розглядають як громадську думку щодо якоїсь політичної проблеми, то є сенс застосовувати критерії її сформованості: наявність емоційно-оцінного ставлення до проблеми, рівень обізнаності з проблемою, наявність програми дій громадськості щодо проблеми, а також потенціал поширення цієї думки в межах певної групи індивідів [4, с. 23–38; 7, с. 23].

У випадку досліджень позицій політичних партій із використанням текстів їхніх програм, перенесення тих самих концептів позиції, які ви-

користуються для дослідження індивідів та груп індивідів, є, на нашу думку, некоректним, щонайменше через дві причини. По-перше, політична партія не є органічною структурою з єдиною «соціально-психологічною диспозицією»; по-друге, текст програми не обов'язково відтворює цінності членів партії і напевно не відтворює цінності *всіх* членів партії. Слід розуміти, що текст є символічною реальністю, яка фіксує певну систему значень, а індивідуальні диспозиції та громадська думка є феноменом радше процесуальним, аніж фіксованим.

Втім, чи відтворює текст програми цінності членів партії, – не є принциповим чинником для використання цього тексту під час аналізу політичних позицій. Адже програма виконує, передусім, функцію стратегічної комунікації з цільовим електоратом, і лише в другу чергу – функцію репрезентації партійців.

Аби запобігти потенційній термінологічній плутанині щодо вживання термінів «політична позиція», «політична установка», «політична орієнтація», ми вживатимемо саме останнє поняття як найближче (але, підкреслимо, *не то-тожне*) до ціннісних орієнтацій індивіда, тобто ідеальних уявлень «як має бути». Зауважимо, що в публікаціях на тему політичних позицій партій терміни «позиція», «установка», «орієнтація», «думка» є, зазвичай, взаємозамінними. Визначення цих понять відбувається на операційному рівні, тому змінюється від дослідження до дослідження. Наприклад, позиція партії щодо зовнішньополітичного вектора визначається тим, чи підтримується в тексті її програми вступ до НАТО, військово-стратегічне співробітництво з Росією, чи й те, й те.

Отже, в цій статті ми розглянемо чотири найпоширеніших методологічних підходи до визначення політичних орієнтацій партій: (1) експертна оцінка партійних орієнтацій; (2) кількісно-якісний контент-аналіз за попередньо розробленими категоріями; (3) словниковий аналіз; (4) шкалювання документів.

Розглядатимемо переваги та недоліки вказаних підходів, базуючись на практиці попередніх досліджень, описаних у літературі.

Експертні оцінки

Експертне оцінювання орієнтацій політичних партій полягає у розміщенні експертами (політичними аналітиками) партій відносно одної вздовж однієї чи більше шкал, які відображають протилежні орієнтації щодо певних проблем. В останні десятиліття цей підхід став одним із найпоширеніших [14; 10, с. 73–88; 16, с. 73–111; 20, 8]. Дослідження свідчать, що експертні оцінки є також і одним із найвалідні-

ших способів визначати партійні орієнтації [19, с. 1–141]. Однак, окрім переваг, цей метод має серйозний недолік: із його допомогою досить складно оцінювати динаміку зміни політичних орієнтацій, оскільки не відомо, які критерії враховують експерти під час оцінки (а якщо відомо – наперед обумовлено – то немає абсолютної певності, що ці критерії експерти інтерпретують однаково), чи змінюються ці критерії, крім того, зміна загальносуспільного (в т. ч. політичного) контексту призводить до врахування інших факторів, які раніше були не відомі. Крім того, складно використовувати національні експертні опитування для міжнародних порівнянь, оскільки експерти імпліцитно враховують місцевий контекст, ранжуючи партії.

Кількісно-якісний контент-аналіз

Цей підхід передбачає розробку кодувальної схеми та використання її експертами-кодувальниками для позначення висловлювань у партійних програмах як таких, що представляють ту чи ту позицію. Найбільш масштабне лонгітудне дослідження з використанням цього методу, як за представленістю країн, так і за часовим охопленням, – «Проект порівняння партійних програм» (Comparative Manifesto Project, CMP). Він аналізує дані щодо 54 країн із 1945 року (українські партійні програми досліджуються з 1994 року). Контент-аналіз програм у рамках цього проекту проводять кодувальники, які маркують «квазі-речення» однією з 56 стандартних категорій [17]. Таким чином можна вимірювати партійні орієнтації з широкого кола проблем: від регулювання ринку до антиглобалізму. Унікальність цього масиву даних полягає в тому, що він дає змогу проводити порівняння як між партіями, так і між різними часовими періодами та між різними країнами (політичними системами), оскільки процедура кодування є стандартною і не залежить від контексту.

Незважаючи на такі очевидні переваги, цей підхід серйозно критикують останніми роками. Зокрема, критики відзначають відсутність індикатора невизначеності, пов'язаної з віднесенням висловлювань до тієї чи тієї орієнтації [12]. Крім того, теоретичний базис цього дослідження, а саме теорія індиферентності партійної конкуренції, зазнає значної критики [9, с. 34]. Відповідно до цієї теорії, політичні партії не конфронтують прямим чином щодо певних проблем, а лише приділяють їм більшу чи меншу увагу. Враховуючи це, CMP вимірює відносну важливість проблем у текстах програм. Дихотомія «праві – ліві орієнтації» визначається через групування висловлювань у «праві» й «ліві» категорії, які є цілком протилежними. Деякі категорії,

живані в рамках дослідження, є біполярними (за/ проти) і тому більш очевидно визначають позиції партій. Інші категорії є монополярними, як-от «націоналізація», що класифікується як «лівий» концепт, чи «закон і порядок», що класифікується як правий концепт. У результаті, шкала «праві – ліві», використовувана у цьому дослідженні, є змішуванням біполярних та монополярних категорій. Проблемою в цьому випадку є те, чи можна прирівнювати вибіркоче висвітлення проблем до партійної позиції. В певних сферах, як-от розподіл економічних ресурсів, моральні проблеми чи інтеграція в міждержавні структури, важко визначити орієнтацію партії лише за даними про обсяг уваги, приділеної цим проблемам у програмах.

Кількісний словниковий аналіз

Цей метод дає змогу автоматично категоризувати частини тексту, використовуючи механічні критерії віднесення. Технічно це реалізується через складання «словника» слів або фраз, кожна з яких прив'язана до певної категорії, після чого комп'ютерна програма автоматично підраховує кількість слів/фраз, що належать до кожної категорії [6, с. 16–18]. Головною перевагою такого підходу є стовідсоткова надійність, оскільки кодувальна схема застосовується щоразу абсолютно однаково, без будь-якої реінтерпретації.

З іншого боку, застосування кількісного словникового аналізу породжує низку проблем. По-перше, на відміну від експертного кодування, ігнорується ширший політичний контекст та змістовний контекст програми в цілому. Одним зі способів подолання наслідків такого контекстуального розриву є внесення до «словника» не лише окремих слів, а словосполучень чи й навіть фраз. Однак це, своєю чергою, суттєво знижує надійність аналізу, бо фрази повторюються набагато рідше, ніж окремі слова. Крім того, поєднання слів у фрази залежить від індивідуального стилю кожного автора тексту. Таким чином, використання окремих слів, які є найменш суперечливими щодо їх належності до тієї чи тієї категорії, є найкращим із можливих у рамках цього підходу рішень.

Хоча в переважній більшості випадків словникове кодування виграє в надійності і програє у валідності, порівняно з експертним кодуванням, існують ситуації, коли кодування, здійснене експертами, є менш валідним, ніж автоматичне комп'ютерне. Знаючи, що програма належить правій партії, експерт може визначити її як праву, не зважаючи на майже однаковий розподіл правих і лівих ідей у ній, чи навіть переважання останніх. Таким чином, словниковий аналіз мо-

же бути більш корисним для визначення неочікуваних змін у політичній риторичі партій, а отже, і в їхніх орієнтаціях.

Автоматичне шкалювання документів

Цей підхід є найновішим у галузі досліджень політичних орієнтацій, що базуються на аналізі політичних текстів. Представлений у 2003 році Левером, Бенуа і Гарі [15, с. 311–331], метод дає змогу автоматично розміщувати тексти на довільних шкалах. Логіка аналізу полягає у такому: відносні частоти слів у текстах (які називають «референтними»), щодо яких відомі їхні позиції у визначених вимірах, порівнюють із відносними частотами слів у текстах, щодо яких позиція не відома («неструктурованих» текстах). Як наслідок порівняння розраховують ймовірність P_{wt} того, що цей текст t пов'язаний із певною позицією, якщо він містить слово w . Тобто, передбачається, що кожне слово надає частку інформації щодо того, який із референтних текстів найбільше відтворює цей неструктурований текст. Оскільки позиції референтних текстів A_{td} відомі, можна використовувати ймовірності P_{wt} разом зі значеннями подібності A_{td} для розрахунку міри S_{wd} для кожного слова w у вимірі d . Потім відносну частоту кожного слова у неструктурованих текстах розраховують як пропорцію F_{wv} відносно загальної кількості слів у тексті. Первинну міру позиції S_{vd} в неструктурованому тексті – як середнє значення S_{wd} всіх слів, які містить цей текст, зважене за частотою всіх прошкальованих слів F_{wv} . Щоби порівняти міри неструктурованих текстів із референтними, проводять трансформацію S_{vd} (докладніше про трансформацію див. [16, с. 326–227]). Після цього розраховують довірчі інтервали через визначення дисперсії V_{vd} між окремими словами відносно середніх значень для кожного тексту.

Перевагами цього підходу, як і у випадку автоматичного словникового кодування, є надійність та можливість порівняння. Крім того, на відміну від словникового аналізу, не створюється жодна апріорна кодувальна схема, натомість валідність залежить від повноти висловлення позиції (якнайширшого лексичного багатства) в референтних текстах та валідності їх віднесення до певної позиції. Для проведення вторинного шкалювання, якщо наявні міри референтних текстів, не потрібно навіть знати мову, якою вони написані. Нарешті, цей підхід передбачає розрахунок міри невизначеності під час віднесення кожного тексту до певної позиції.

Недоліком автоматичного шкалювання є, по-перше, обмеженість лексичної повноти навіть у найоб'ємніших програмах. По-друге, референтні й неструктуровані тексти мають бути по-

дібними за стилістикою. По-третє, трансформація, яку використовують для точнішого визначення значень неструктурованих текстів по шкалі, залежить від комбінації цих текстів (трансформація передбачає відбір трьох довільних неструктурованих текстів із загальної вибірки). Для подолання цієї проблеми було запропоновано інший варіант трансформації, що передбачає розміщення всіх текстів (референтних і неструктурованих) на одній шкалі з використанням відносних дистанцій згенерованих первинним розрахунком мір позицій. Втім, сьогодні тривають суперечки щодо коректності першого і другого варіанта трансформації.

Порівняння підходів. Висновок

У таблиці нижче ми наводимо найважливіші характеристики розглянутих підходів.

Як бачимо, найбільшу кількість переваг мають експертні опитування і новітня методика шкалювання текстів. Отже, для досягнення найкращого результату було б доцільно комбінувати повністю автоматичний аналіз текстів з експертними оцінками значення тих чи тих слів і комбінацій слів. Одним із варіантів розв'язання цієї проблеми є алгоритм виявлення змістовних зв'язків між різними концептами в межах одного тексту, що його ми запропонували в іншій статті [2, с. 335–340].

Наостанок зазначимо, що описані підходи можуть застосовуватись не лише для визначення політичних орієнтацій, а й у будь-якій іншій сфері, де є доступні тексти з максимально вичерпним представленням позицій опитуваних (наприклад у напівструктурованих глибинних інтерв'ю).

Таблиця. Підходи до визначення політичних орієнтацій на основі текстових даних (зведені характеристики)

Підхід	Перевага валідності/ надійності	Необхідність попередньої трансформації інструментарію для конкретної задачі	Необхідність попередньої обробки текстів	Мінімальна кількість людей, необхідних для проведення аналізу	Можливість міжнародних порівнянь	Можливість часових порівнянь
Експертні оцінки	Валідність	Є. Підбір експертів з певної проблеми	Немає	Як правило, не менше 3 для проведення перевірки узгодженості	Ускладнена: потребує створення міжнародної групи експертів та уніфікації критеріїв	Ускладнена: потребує створення процедур контролю незмінності критеріїв оцінки
Кількісно-якісний контент-аналіз	Надійність і валідність залежать від якості кодувальної схеми і підготовки кодувальників. Немає переважання	Є	Немає	3		
Кількісний словниковий аналіз	Надійність	Є	Немає	1		
Автоматичне шкалювання	Надійність	Немає	Є. Необхідний відбір тих частин програми, які стосуються саме аналізованої проблеми. Необхідний попередній відбір і маркування референтних текстів	1		

1. Аспекти зовнішньополітичної діяльності політичних партій на шляху до євроатлантичної інтеграції України : матеріали всеукраїнської науково-практичної конференції. [«Україна-НАТО: регіональний вимір»], (Львів, 6 грудня 2008 року) / [укл. А. С. Романюк, Л. Скочиляс]. – Львів, 2008. – С. 16–19.
2. Іванов О. В. Комп'ютерний контент-аналіз: проблеми та перспективи вирішення / Олег Валерійович Іванов // Методологія, теорія та практика соціологічного аналізу сучасного суспільства. – 2009. – № 15. – С. 335–340.
3. Контент-аналіз виборчих програм політичних партій та блоків / Український незалежний центр політичних досліджень // Пріоритети. – 2006. – № 3 (29). – С. 13–25.
4. Оссовський В. Л. Соціологія громадської думки / Володимир Леонідович Оссовський. – К. : ПЦ «Фоліант» ; Вид. дім «Стилос», 2005. – С. 23–38.
5. Саморегуляція и прогнозирование социального поведения личности / [под ред. В. А. Ядова]. – Л. : Наука, 1979. – С. 47–62.

6. Alexa M. Computer-assisted text analysis methodology in the social sciences / Melina Alexa // ZUMA-Arbeitsbericht. – 1997. – V. 7. – P. 16–35.
7. Allport F. H. Toward a Science of Public Opinion / F. H. Allport // Public Opinion Quarterly. – 1937. – V. 1. – P. 23.
8. Benoit K. Party Policy in Modern Democracies / Kenneth Benoit, Michael Laver. – London : Routledge, 2006. – 304 p.
9. Budge I. Explaining and predicting elections. Issue Effect and Party Strategies in Twenty-Three Democracies / I. Budge, D. J. Farlie. – London : Allen and Unwin, 1983. – P. 34.
10. Castles F. Left-right political scales – some expert judgement / F. Castles, P. Mair // European Journal of Political Research. – 1984. – V. 12 (1). – P. 73–88.
11. Clinton J. D. Integrating Voting Theory and Roll Call Analysis: A Framework / Joshua D. Clinton, Adam Meirowitz // Political Analysis. – 2003. – V. 11 (4). – P. 381–396.
12. Fickle parties of changing dimensions? Testing the Comparability of the Party Manifesto Data Across Time and Space: Paper presented at the MPSA conference, (Chicago, 2006) / R. Bakker, E. Edwards, C. de Vries. – Chicago, 2006. – P. 23–35.
13. Huber J. D. Expert interpretations of party space and party locations in 42 countries / J. D. Huber, R. Inglehart // Party Politics. – 1995. – V. 1 (1). – P. 73–111.
14. Janda K. Political Parties. A Cross-national Survey / Kenneth Janda. – N. Y. : Free Press, 1980. – 960 p.
15. Laver M. Extracting policy positions from political texts using word as data / Michael Laver, Kenneth Benoit, John Garry // American Political Science Review. – 2003. – V. 97 (2). – P. 311–331.
16. Laver M. Policy and Party Competition / Michael Laver, W. Ben Hunt. – N. Y. : Routledge, 1992. – 256 p.
17. Mapping Policy Preferences: Parties, Electors and Governments: 1945–1998 / I. Budge, H.-D. Klingemann, A. Volkens, J. Bara, E. Tannenbaum (Eds.) – Oxford : Oxford University Press, 2001. – 388 p.
18. Marks G. Crossvalidating data on party positioning on European integration / Gary Marks, Liesbet Hoogh, Marco R. Steenbergen, Bakker Ryan // Electoral Studies. – 2007. – V. 26 (1). – P. 23–38.
19. Marks G. Estimating error in measures of party positioning: expert, manifesto, and survey data compared / Gary Marks // Electoral Studies. – 2007. – V. 26 (2). – P. 1–141.
20. Marks G. Chapel Hill Party Dataset 2002 / G. Marks, M. Steenbergen. – Chapel Hill : University of North Carolina, 2002. – 320 p.

O. Ivanov

ESTIMATING POSITIONS USING UNSTRUCTURED TEXTUAL DATA: APPROACHES EVALUATION

In the article four the most popular modern approaches to the positions estimation in a given sphere with the use of unstructured textual data analysis were presented: expert surveys, qualitative content-analysis, quantitative dictionary-based analysis and document scaling without pre-constructed coding scheme. The evaluation and comparison of the approaches were done concerning their validity and reliability, technical convenience, effectiveness and resource intensiveness. All four approaches were used for the analysis of the foreign policy positions of Ukrainian parties, as presented in their election manifestos.

Keywords: *content-analysis, politics, policy positions, natural language processing, parties, manifestos, expert surveys.*