

Бондар В. С., Допіра М. А.

РОЗГЛЯД МЕТОДУ КОНТЕНТ-АНАЛІЗУ З ПОГЛЯДУ КІЛЬКІСНО-ЯКІСНИХ ТЕХНІК ПРОВЕДЕННЯ

У статті зроблено спробу проаналізувати контент-аналіз як кількісно-якісний метод соціологічного дослідження. Розглянуто найбільш поширені техніки, логіку їхнього аналізу, сфери застосування та можливості.

Якісно-кількісний аналіз документів на теренах Радянського Союзу з'явився ще на початку 70-х років ХХ ст. [1] і переважно застосовувався як допоміжна техніка дослідження. При цьому основна увага приділялася підрахунку частот тих чи інших символічних елементів у тексті. З початку 90-х років зацікавленість у методах аналізу документів, і методи контент-аналізу зокрема, суттєво зросла, адже подібні відносно недорогі інструменти дослідження дозволяють виявити зміни, що відбуваються в суспільстві, виділяючи прихований зміст повідомлень. Проте через брак якісної вітчизняної наукової літератури стосовно методології проведення аналізу документів і відсутності належного досвіду здійснення подібних досліджень дослідники-практики часто не мають чіткого уявлення про сутність методу контент-аналізу та його різноманітних технік. Науковці не приходять до спільної думки, які техніки можна віднести до даного методу, якою мірою цей інструмент є кількісним, а якою – якісним. Тож дана роботи полягає в намаганні співставити різні погляди на сутність контент-аналізу як методу та його різноманітні техніки. На сьогодні не існує узагальненої наукової роботи, яка містила б співставлення різних думок дослідників стосовно дослідницьких технік, що загалом називаються контент-аналізом, а останнім часом ще й іноземним виразом Data Mining, що означає видобування даних і виокремлення в них прихованого змісту, а також стислого викладу технік, які застосовуються при проведенні контент-аналізу, опису сфери їхнього використання, сильних і слабких сторін. Виходячи з вищезазначеного, ця спроба може стати в нагоді як досліднику-практику, котрий вирішує питання вибору найбільш доцільного методу аналізу, так і теоретику, що цікавиться проблемою співвідношення кількісних і якісних компонентів контент-аналітичного дослідження. Тож об'єктом дослідження є контент-аналіз як метод соціологічного дослідження,

а предметом – специфіка контент-аналізу як кількісного-якісного методу аналізу документів із різноманітними дослідницькими техніками. Мета роботи полягає в розгляді контент-аналізу як багатогранного кількісного-якісного методу, сутність якого по-різному трактується дослідниками.

Сучасне суспільство продукує безліч матеріальних та інформаційних одиниць, котрі поширюються, споживаються та відтворюються в соціальній структурі. Дослідження цих одиниць є надзвичайно цінним для розуміння явищ, що мають місце в суспільстві, а також змін, які відбулися або відбуваються. Саме ці завдання ставить перед собою дослідник при використанні аналізу текстів або інших документів.

Аналіз тексту, або в ширшому розумінні аналіз документів, може здійснюватися багатьма способами, що включають як кількісні, так і якісні методи. Кожен із них має особливий підхід до аналізу, інструментарій та сферу застосування. Утім велика кількість таких способів має єдину назву – «контент-аналіз». Кожен дослідник визначає цей метод відповідно до власних поглядів і вподобань. У науковому середовищі не закріпилося єдиного визначення, яке б окреслювало межі даного дослідницького методу. Для того, щоб зрозуміти, що вкладається в словосполучення «контент-аналіз», а також у більш загальне поняття «аналіз документів», варто звернутися до наявних визначень.

Історично першим виник аналіз документів, що сьогодні називається традиційними. Мета такого аналізу полягає в інтерпретації тексту в цілях дослідника. За допомогою цього методу дослідник намагається дістатися сутності документа; виявити мотиви та задуми комунікатора; умови, за яких документ був створений. Таким чином, аналіз включає як внутрішній, так і зовнішній аспекти розгляду. Основною метою аналізу є формування загального уявлення про документ,

оцінка його відповідно до завдань дослідження. Проте, на відміну від простого ознайомлення, традиційний аналіз ставить на меті дійти до суті документа, виявити його внутрішню логіку, достовірність, обґрунтованість, образно кажучи, — перетравити його [2]. Очевидно, що подібний аналіз цілком залежить від установок дослідника і є суб'єктивним.

Як альтернатива такого підходу до аналізу документів виник метод формалізованого аналізу змісту, що дістав назву контент-аналізу. Класичне визначення Берельсона таке: «Контент-аналіз — це дослідницька техніка для об'єктивного, системного й кількісного опису наявного змісту комунікації, яка відповідає цілям дослідника» [3]. Деякі дослідники, наприклад В. Ядов, наголошують на кількісному аспекті контент-аналізу, пропонуючи таке визначення: «Контент-аналіз — це переведення в кількісні показники великої кількості текстової (чи записаної на плівку) інформації» [2]. М. Розов, на противагу, пропонує методiku цілком традиційного внутрішнього аналізу, побудовану на логічній конструкції тексту, що також називає контент-аналізом [4]. Його метод використовується для розгляду завершених філософських текстів і слугує для визначення суб'єктно-об'єктних взаємодій, логічної конструкції, онтології, аксіології тексту та взаємодій його частин. При цьому дослідник наголошує на тому, що висновки, отримані в результаті подібного аналізу, будуть відрізнятися залежно від людини, що проводить процедуру, а сам процес є більше творчим, аніж механічним.

Осгуд визначає контент-аналіз як «процедуру, за допомогою якої можна зробити припущення щодо джерела інформації та адресата, виходячи з повідомлення, яким вони обмінюються» [3]. Розглянувши понад 40 визначень контент-аналізу, М. Костенко та В. Іванов зробили спробу дати узагальнене визначення контент-аналізу, що виглядає так: «Контент-аналіз — це якісно-кількісний метод вивчення документів, який характеризується об'єктивністю висновків і строгістю процедури та полягає у квантифікаційній обробці тексту з подальшою інтерпретацією результатів. Предметом контент-аналізу можуть бути як проблеми соціальної дійсності, котрі висловлюються чи, навпаки, приховуються в документах, так і внутрішні закономірності самого об'єкта дослідження» [1].

Для того, щоб не розгубитися у величезній кількості визначень, Робертс пропонує доволі цікаву модель класифікації визначень контент-аналізу залежно від їхнього спрямування [3]. По-перше, визначення різняться залежно від того, які

об'єкти включаються в аналіз — текст чи інший символічний матеріал (як-от комунікація, символічна поведінка, знак-переносник, повідомлення тощо). По-друге, визначення мають розбіжності стосовно того, який інтелектуальний продукт може бути отриманий в результаті процедури контент-аналізу — опис, висновки чи класифікації. По-третє, немає однозначності в питанні, чи може (і чи мусить) контент-аналіз виходити за межі явного змісту, чи повинен також розглядатися латентний зміст. По-четверте, дослідники не однастайні в питанні, стосовно чого можна використовувати висновки, отримані в результаті контент-аналізу — стосовно тексту (чи іншого символічного матеріалу), джерела інформації чи одержувачів (аудиторії). Ще одне джерело неоднозначності міститься у визначенні характеру контент-аналізу — він вважається або кількісним, або якісним, або ж змішаним методом. Щодо останнього існують розбіжності в тому, якою мірою перша та друга складові притаманні контенту.

Варто зазначити, що багато визначень не містять однозначної позиції стосовно кожного пункту розрізнення, проте кожен дослідник повинен для себе визначити, що для нього є контент-аналіз за всіма цими пунктами перед тим, як проводити дослідження. Якщо спробувати сформулювати найбільш коротке визначення, що адекватно відбиватиме сутність контент-аналізу, то твердження, з яким важко сперечатися, виглядатиме приблизно так: *контент-аналіз є вимірюванням тексту чи іншого символічного матеріалу, що має методологічне підґрунтя й відповідає завданням соціальних наук*. При цьому контент-аналіз має *систематично редукувати текст* (потік повідомлень чи інших символів) до визначеного набору символів, який можна піддавати статистичній обробці. Саме можливість робити узагальнені висновки стосовно джерела, адресата чи самого повідомлення і є тією рисою, що відрізняє контент-аналіз від традиційного аналізу документів.

Основоположники контент-аналізу Берельсон і Картрайт вважали цей метод суто кількісним. Ці дослідники передбачали, що продукт, одержаний у результаті аналізу, повинен мати форму опису, причому опису за допомогою чисел [3, 27]. Із розвитком методу про нього стали згадувати радше як про кількісно-якісний. Якісні компоненти вбачалися у свободі дослідника визначати категорії аналізу та підрахунку [5].

Інші дослідники вважають контент-аналіз якісним методом завдяки розробці різноманітних класифікацій і виявленню тих чи інших структурних закономірностей у документах [6]. На

думку деяких учених [1, 6], якісний компонент контент-аналізу полягає саме у визначенні категорій, а разом із ними й одиниць аналізу, що цілком залежать від намірів дослідника.

Зазвичай прихильники кількісної парадигми проведення контент-аналізу наголошують, що дану дослідницьку техніку є сенс використовувати для великих масивів інформації, причому бажано однопланової та неунікальної [2]. У випадку, коли йдеться про унікальні чи одиничні матеріали, контент-аналіз є неналежним методом дослідження. Проте прихильники більш широкого кількісно-якісного трактування контент-аналізу зазначають, що якісний характер даної техніки дозволяє робити висновки не тільки з наявних у документі категорій, а й із відсутніх [1, 7]. Контент-аналіз, побудований на якісній методології, має декілька цікавих напрямів, зокрема виявлення обмовок, описок, неточностей у документі, що начебто є випадковими. Проте, якщо вірити Фройдю, подібні неточності відображають підсвідомі наміри людини. І якісний аналіз таких несвідомих помилок може бути базою для цікавих висновків дослідника.

Варто зазначити, що якісний контент-аналіз відрізняється від традиційного аналізу документів. Замість реферування тексту якісний контент-аналіз має на меті виявити основні теми, що зустрічаються у тексті. Однак проблемою може стати велика кількість текстів для аналізу. Якщо вибірка є аж надто великою, на допомогу приходять комп'ютеризовані засоби обробки інформації. На основі тезаурусу, що в ідеалі повинен вмещувати та класифікувати всі наявні в документі слова, будуються частотні графіки для кожної категорії. Як бачимо, процедура дуже нагадує звичайний кількісний контент-аналіз. Проте дослідник, що користується якісною парадигмою, звертатиме увагу не тільки на категорії, що зустрічаються найчастіше, а й на ті, які майже не з'являються, але присутні в тексті. Такими категоріями можуть бути вже згадані обмовки людини чи, наприклад, категорії скрути, непридатності, безвихідності у прес-релізах успішної компанії тощо [8]. Джерелом цікавих висновків для дослідника можуть бути не тільки теми, що зустрічаються в тексті, а й ті, що взагалі не фігурують у ньому. Наприклад, якщо свідомо оминається тема, яка за логікою повинна бути присутня в документі, це наводить на певні роздуми.

Ще однією інтерпретацією якісного контент-аналізу є думка, що такий метод дослідження має стосуватися не тільки й не стільки змісту документа, скільки його оформлення, розташування серед інших матеріалів, шрифту, наявності

ілюстрацій чи малюнків тощо [7]. При цьому, як було вже зазначено, зміст документа розглядається тільки поверхово.

Контент-аналіз часто розглядають як метод аналізу даних, одержаний у результаті якісних досліджень. Найчастіше такими «сирими» даними є транскрипти інтерв'ю чи фокус-груп. При цьому в дослідника часто немає жорсткої схеми, набору категорій, що він виявляє. Натомість він фіксує появу тих чи інших одиниць, що здалися цікавими й релевантними [9]. Подібний підхід різко контрастує із традиційною методикою контент-аналізу, де окреслення категорій та одиниць аналізу є обов'язковим попереднім кроком.

На жаль, навіть сучасні дослідники визначають контент-аналіз як виключно кількісну техніку, що зводиться до «підрахунку ознак, рис, властивостей документа (тексту)» [10]. То м у є сенс зупинитися на висвітленні основних технік кількісно-якісного аналізу, що на даний момент набули найбільшого поширення серед науковців і можуть бути виконані сучасним програмним забезпеченням.

За допомогою сучасних програм, створених для контент-аналізу, виділяють такі основні можливості для аналізу текстових документів:

1. Знаходження слова чи комбінації слів у документі чи виділеному фрагменті. Така процедура доступна в більшості текстових редакторів.

2. Підрахунок частот появи слів у документі чи сукупності документів. Часто в цю процедуру входить порівняння частот появи в різних документах.

3. Стандартизація мовних одиниць. Ця процедура покликана стандартизувати текстові одиниці задля уможливлення порівнянь. Конкретними методиками для досягнення даного результату є:

- а) визначення походження слів із огляду на їхню побудову (префікси, суфікси, інфікси тощо);

- б) згуртування слів у випадках, коли декілька слів є ідіоматичною конструкцією чи географічною назвою і є сенс рахувати таке словосполучення як одне слово;

- в) усунення двозначностей. У випадках, коли слово має більше, ніж одне значення, за контекстом обирається найбільш доречне.

4. Категоризація одиниць. Дана процедура розбиває текст на категорії, визначені дослідником відповідно до тезаурусів, побудованих чи заданих автоматично. У контексті процедури підраховується кількість одиниць, що відповідають тій чи іншій категорії, і здійснюється їхнє порівняння.

5. Пошук комбінацій мовних елементів і наявних у тексті сталих сполучень.

6. Кластеризація тексту відповідно до автоматично виділених категорій.

7. Виявлення закономірностей у документі чи сукупності документів.

8. Співставлення текстових сегментів для виявлення подібних чи протилежних тенденцій.

9. Побудова семантичних мереж тексту, що визначають найбільш часто вживані слова та їхні сполучення.

10. Побудова тематичних дерев (що є видозміненими мережами), які висвітлюють не тільки структуру, а й частоту згадувань тих чи інших понять (вузлів) мережі, тобто їхню значимість.

11. Реферування тексту на основі тематичного дерева, коли в тексті виділяються речення (словосполучення), що містять у собі найбільшу кількість значимих елементів (категорій).

12. Виділення емоційно забарвленої лексики, співставлення її з певними категоріями, персоналіями тощо.

13. Знаходження слів чи словосполучень, що є іменниками, які позначають власні назви: торгових марок, організацій, географічних об'єктів тощо.

14. Використання нетекстової інформації в поєднанні з текстовим аналізом. Ця процедура передбачає включення демографічної, архівної чи іншої нетекстової інформації в статистичний аналіз.

Деякі із цих процедур є доступними для користувачів стандартних офісних програм, таких як Office чи Adobe Reader. Одні з них передують аналізу, інші складають його основу. Саме на цих, останніх у списку основоположних для процедури аналізу, техніках і зосередимо увагу. Їхній розгляд відбуватиметься за схемою: логіка аналізу та його процедура; сфери застосування техніки; сильні сторони методу; слабкі сторони методу; загальна оцінка.

Підрахунок частот

Найдавнішою та найпоширенішою процедурою контент-аналізу є підрахунок частот тих чи інших елементів тексту, що беруться як одиниці аналізу й діляться на дві великі групи: змістовні та граматичні. Перші відтворюють зміст повідомлення, другі — його композицію та певні глибинні установки автора. Для успішного проведення підрахунку граматичних одиниць документа потрібно дотримуватись повної теорії, що обґрунтовує зв'язок форми подачі думок із задумами автора. Наприклад, виходячи з того, що прийменник «до» (російською «к») відповідає спрямованості в майбутнє, а прийменник «від» свідчить про орієнтацію на минуле, можна дослідити

промови політичних лідерів і прийти до цікавих висновків [11]. Іншим прикладом слугує підрахунок слів і словосполучень, що виражають невпевненість, таких як «мабуть», «можливо», «на мою думку» тощо. Попередньо, звичайно, потрібно створити тезаурус слів, що будуть підраховуватися. Можна спробувати укласти словники слів, що не несуть змістовного навантаження, але які вживають політики в різних ситуаціях. Таким чином може бути проведено дослідження прихованих мотивів, або когнітивної складності сприйняття описуваного феномена автором, або ж зміни його поглядів.

Утім подібні дослідження є нечисленними, складними для обґрунтування валідності й у цілому нетиповими для соціології. Набагато більше уваги прикуто до змістового підрахунку частот, який є найпоширенішою технікою контент-аналітичного дослідження.

Найпростішою формою такого дослідження є так звана бінарна оцінка, тобто визначення того, чи даний текст або його уривок стосується певної теми, чи згадуються в них певні події тощо. Важливо зазначити, що фіксується тільки факт наявності категорії аналізу, яка цікавить дослідника, при цьому кількість згадувань не враховується. Результатом такого аналізу є співвідношення кількості текстів, що містять необхідні категорії, до всієї сукупності. Розвитком цієї техніки аналізу є власне частотний аналіз, що нині є більш популярним при вирішенні завдань підрахунку інтенсивності висвітлення категорій.

Логіка та процедура

За логікою цієї процедури першим кроком аналізу є виявлення категорії аналізу — «міського», разом із тим лаконічного вираження проблеми», за влучним визначенням Федотової [12]. Категорія не є безпосередньо даною в тексті у вигляді, сформульованому дослідником, вона радше є характеристикою всієї сукупності текстів. Це достатньо абстрактне поняття, на яке екстраполюються результати дослідження. Подібним поняттям можуть виступати, наприклад, «расова дискримінація», «оперативність подання інформації», «прихильність до певної політичної сили», «географічна збалансованість подачі інформації» тощо. Виокремлення категорій повинно відбуватися водночас із постановкою проблеми дослідження й адекватно відбивати її. Визначення категорій є універсальним кроком, що потрібний при використанні будь-якої методики аналізу, у тому числі й для підрахунку частот.

Після визначення категорій потрібно конструювати категорійну сітку, тобто класифікувати та згрупувати семантичний простір аналізу відповідно до завдань дослідження. Іншими словами, різні модифікації категорій аналізу слугуватимуть опорними моментами для відбору потрібної інформації, будуть своєрідним решетом, що затримає потрібні елементи тексту.

Наступним кроком є визначення одиниць аналізу. Ними можуть бути: слова, речення, теми, ідеї, автори, персонажі, соціальні ситуації чи інші частини тексту, об'єднані чимось, що відповідає змісту контент-аналізу [13]. Одиницею аналізу можуть також слугувати окремі матеріали чи повідомлення.

Фінальним кроком операціоналізації понять є вибір одиниць підрахунку – кількість слів чи словосполучень, рядків, друкованих знаків, сторінок, абзаців, довжина повідомлень, площа тексту чи малюнків, виражена у фізичних одиницях, тощо.

При виділенні одиниць аналізу та підрахунку використовуються різноманітні методи, які можна поділити на дві великі групи:

1) одиницею аналізу є ознака матеріалу, що характеризує його в цілому. Це може бути оперативність, характер матеріалу, його розміщення тощо. Підрахунок подібних одиниць повинен відбуватися відповідно до розробленої дослідником класифікації, що включає в себе всі можливі варіанти. Наприклад, якщо аналізується тематика виступу чи публікації, то всі варіанти мають бути продуманими. У деяких випадках це неможливо чи недоцільно, тому в такому разі додаються пункти на кшталт «матеріали, що не можуть бути класифіковані» або «інші»;

2) одиницею аналізу може виступати окремий фрагмент документа, що стосується тієї чи іншої теми. Він може бути фіксований як жорстко, так і нежорстко. Проте в усіх випадках ідеться про достатньо обмежену кількість слів, суджень чи інших елементів, що цікавлять дослідника. При цьому: підраховуватися може й частота появи певного слова, і кількість речень, абзаців тощо, в котрих зустрічається це слово, і кількість матеріалів, у яких принаймні раз фігурувало це слово. Подібну процедуру можна відтворити й при підрахунку площ матеріалів, де зустрічалось те чи інше слово або елемент.

При проведенні аналізу частотного розподілу одиниць найскладнішим етапом дослідження є визначення категорій аналізу. Через те, що дані елементи визначаються дослідником і є основоположними для всього аналізу, їхній правильний вибір є надзвичайно важливим для успішного

проведення всього аналізу. Категорії не повинні бути як надто абстрактними, щоб їх можна було об'єктивно дослідити, так і надто специфічними, щоб висновки, одержані в результаті контент-аналізу, можна було поширити на певне явище соціального життя, що виходить за межі документа, який досліджується. Утім існують певні загальні правила вибору категорій, що дозволяють адекватно провести контент-аналіз. Згідно з Григор'євим [13], категорії аналізу повинні бути:

• доречними (тобто валідними для дослідження визначених феноменів);

• вичерпними, взаємовиключними (тобто такими, щоб одна одиниця аналізу чи підрахунку не входила до складу різних категорій в однаковому обсязі);

• надійними (такими, щоб класифікація була прозорою й однозначною).

Ще одним цікавим різновидом частотного аналізу є підрахунок оцінок і пропозиційних елементів, таких як «Наші законодавці повинні зрозуміти...», «Без дотацій фермерство не може звестись на ноги». У цілому можна підраховувати будь-які оцінки, поділяючи їх на позитивні та негативні, і на цьому ґрунті робити висновки. Одиницями підрахунку також можуть слугувати метафори з їхнім певним забарвленням, інші поетичні елементи, складні понятійні конструкції тощо.

Проте в усіх описаних випадках основним результатом аналізу є отримання частоти появи тієї чи іншої одиниці аналізу в документі та порівняння цього числа з іншими текстами у сукупності.

Особливості аналізу

Крім простого підрахунку кількості появи тієї чи іншої категорії аналізу, представлені як сума появ одиниць аналізу, соціологи використовують декілька методик, що дозволяють отримувати значно цікавіші дані. По-перше, це підрахунок відносних частот, що вираховуються як відношення абсолютної частоти до довжини аналізованого матеріалу. При цьому довжина може обчислюватися кількістю символів, слів, речень, абзаців тощо.

У загальному вигляді формула коефіцієнта Яніса, призначена для підрахування відношення позитивних і негативних оцінок, суджень, аргументів і т. д., у випадку, коли позитивних оцінок більше, ніж негативних, виглядає так:

де c – коефіцієнт Яніса, f – число позитивних оцінок, n – число негативних оцінок, r – обсяг тексту, що має відношення до проблеми, яка вивчається, t – загальний обсяг текстів.

У випадках, коли кількість негативних оцінок перевищує кількість позитивних, формула набуває такого вигляду:

$$c = \frac{f \cdot n - n^2}{r \cdot t}$$

І слугує для виявлення різниці позитивних і негативних оцінок. При цьому, якщо переважають негативні оцінки – коефіцієнт є від'ємним, якщо ж превалюють позитивні – додатним.

Сфери застосування та умови використання

Передусім частотний розподіл даватиме адекватні дані в тому випадку, коли є сукупність однопорядкових даних. При цьому для дослідження важливим є точність підрахунку одиниць. Зазвичай цей метод використовується тоді, коли матеріалу не просто багато, а настільки багато, що його не можна охопити без сумарних оцінок [8]. Важливою передумовою є достатня кількість згадування досліджуваних якісних характеристик у тексті. Утім деякі дослідники [4; 14] наголошують, що треба зважати не тільки на наявні категорії та одиниці аналізу, а й на ті, що відсутні в документі.

Обмеження у використанні та слабкі сторони

Першим застереженням при використанні кількісного контент-аналізу, що має на меті виявити частоти появ тих чи інших категорій, є випускання з процесу аналізу цілей, які ставили перед собою автори аналізованих документів. Навіть при удаваній одноплановості текстів, вони могли бути створені з різними цілями. Тільки враховуючи специфіку створення об'єктів аналізу, можна одержати адекватні результати. Виходячи з визначення контент-аналізу як систематичного редукування тексту, одним із основних недоліків кількісної інтерпретації текстів є випускання багатьох елементів, які, можливо, є цінними для дослідження. До того ж різні документи були створені для різної аудиторії. Різний склад людей, що сприйматимуть текст чи інше послання, безумовно вплине на характер створеного документа. Ще одним важливим обмеженням слугує недоступність матеріалів, які становлять генеральну сукупність документів, що стосуються теми дослідження. Частотний розподіл має сенс використовувати на достатньо великих за обсягом вибірках для того, щоб результати до-

слідження могли поширюватися на всю сукупність. Загалом вважається, що вибірка для проведення кількісного частотного аналізу повинна бути 200–600 текстів [9]. Вибір окремих текстів повинен бути обґрунтованим і максимально випадковим. Проте в дійсності дослідник не має в розпорядженні всього обсягу матеріалів, тож йому доводиться задовольнятися існуючими наявними, що суттєво погіршує метод.

Наступною перепоною для ефективного проведення частотного аналізу є повнота тезаурусів, тобто списків слів, що стосуються однієї категорії чи одиниці аналізу; списків синонімів, якими може бути замінене слово; списків позитивно та негативно забарвлених слів, що стосуються певної теми тощо. Особливої гостроти ця проблема набуває при машинному аналізі, адже людина може з більшою легкістю змінювати свій тезаурус і включати до нього нові слова та вирази. Щодо англійської мови, то різноманітні тезауруси існують уже давно й дають змогу процес аналізу зробити максимально стандартизованим і незалежним від дослідника. У російській мові перший семантичний словник РУСЛАН поки що перебуває на стадії розробки [11]. Прецедентів ґрунтового вивчення семантичного поля української мови не знайдено. Виходячи з такої ситуації, кожен дослідник повинен самостійно створювати тезаурус, на який орієнтуються кодувальники, або ж користуватися тезаурусами комп'ютерних програм, що автоматизують процес контент-аналізу. Очевидно, всі тезауруси різнитимуться між собою, що призводить до значного зменшення об'єктивності отриманих даних. Проте дана проблема характерна не тільки для досліджень неангломовних джерел. Робертс [3] застерігає дослідників від довіри комп'ютерним програмам, що можуть «робити помилки у визначенні теми там, де цього не зробить семирічна дитина».

Ще одним недоліком слугує неможливість дослідити явище, що цікавить дослідника, у повній мірі. Він повинен робити певні висновки на основі наявної інформації, яка рідко є релевантною на сто процентів.

Загальна оцінка

Підрахунок частот є базовим видом аналізу, що дозволяє побачити структуру тексту, його акценти й одразу виявити закономірності. Також дана техніка є підготовчою для багатьох складніших процедур дослідження. Будучи найпростішою й не позбавленою серйозних недоліків, ця техніка все ж таки дозволяє приходити до цікавих висновків.

Дослідження емоційної забарвленості, відношення автора

Логіка аналізу

Інший метод, що знаходить широке застосування при проведенні контент-аналітичного дослідження, є виявлення емоційної забарвленості документа, ставлення автора до висвітлюваної теми. Подібні дослідження базуються на явищі синестезії, тобто переносу категорій із одної сфери в іншу. Наприклад, такі вирази, як «тверде переконання», «теплі стосунки», «запізнілий виступ» тощо, мають у собі певну оцінку, яку можна виявити; можливо, навіть певну мотивацію до дій.

Часто при проведенні подібного дослідження користуються вимірами семантичного диференціалу. Цей метод був запропонований американським психологом Осгудом у середині минулого сторіччя [15]. Дана дослідницька техніка базується на парах прикметників-антонімів, між якими є неперервна шкала проміжних положень. Оцінюючи дещо, в нашому випадку текст, дослідник може характеризувати його за численними категоріями, проте базовим набором факторів, на які варто орієнтуватися, є сила, активність і оцінка [11].

Наступним кроком є підрахунок прикметників, що мають певне емоційне забарвлення, й оцінка інтенсивності цього забарвлення. Для цього в нагоді стануть комп'ютерні програми аналізу тексту, в які вмонтовано лексикони емоційно забарвлених слів, проранжованих за інтенсивністю забарвлення. Іншим способом підрахунку є ручний. У цьому разі кодувальник виставлятиме оцінку емоційній забарвленості елементів за 9-бальною шкалою. За методикою в нього будуть квоти на кожен із ступенів забарвленості, що відповідають нормальному розподілу. Інший підхід до цієї ж процедури полягає в порівнянні всіх наявних елементів (а ними можуть бути як слова, так і тексти) між собою для визначення більш чи менш емоційно забарвлених.

Сфера використання, сильні та слабкі сторони

Даний метод найчастіше використовується для дослідження публічних промов, виступів, публіцистичних матеріалів і т. п. Не варто його застосовувати до наукових праць, звітів чи прес-релізів. Особливо ефективним є використання даного методу для дослідження джерел, що позиціонують себе як незалежні та незаангажовані, тому що за допомогою саме цієї техніки можна

перевірити правильність подібних тверджень. Утім метод може давати тривіальні результати і бути неефективним для розгляду дуже емоційних текстів.

Загальна оцінка

Звичайно, лексика легше контролюється автором документа порівняно з граматикою. Глибоке семантично-лінгвістичне дослідження граматичної будови тексту може більше сказати про ставлення автора до описаної проблеми. Проте такі методи зазвичай передбачають, як уже було зазначено, певну теорію й певний набір знань у сфері, де соціолог не є компетентним.

Семантична мережа

Логіка та процедура

Кожний документ має певну логічну побудову. Предмети та події, описані чи показані, повинні певним чином взаємодіяти між собою, утворювати деяку цілісну мережу, яка є згорнутим змістом документа. Деякі поняття в тексті чи сукупності текстів мають більшу ймовірність бути поєднаними, ніж інші. Наприклад, цілком очевидно, що поняття «Верховна Рада» і «депутат» скоріше за все матимуть сильний взаємозв'язок, а «депутат» і «черевики», якщо, звісно, у тексті згадувалися черевики, будуть непов'язаними чи принаймні слабо пов'язаними. Дослідження семантичних мереж дозволяє виявити типові поєднання понять, оточення певного обраного об'єкта тощо.

Механізм виявлення зв'язку побудований на припущенні, що явним показником зв'язку між предметами чи явищами є опис їхньої участі в одній події (незалежно від виконуваної ролі). При цьому показником наявності зв'язку між предметом та подією є опис участі предмета в даній події (незалежно від виконуваної ролі) [16]. У загальному вигляді семантична структура тексту складається з так званих трійок (поєднань об'єкт – зв'язок – об'єкт).

У семантичну мережу зазвичай включаються найуживаніші слова тексту, що несуть смислове навантаження. Таким чином, побудова мережі є неможливою без попередньої побудови частотного розподілу.

Побудова семантичної мережі передбачає використання лінгвістичних алгоритмів морфологічного та синтаксичного аналізу, а також тезаурусу мови. При цьому на попередньому етапі здійснюють (зазвичай машинну) підготовку тексту до аналізу. Процедури підготовки передбачають зведення синонімів і синонімічних виразів до одного поняття, а з числа категорій

виключаються загальноживані слова та слова, що не несуть самостійного смислового навантаження. Наприклад, слова «концепція» чи «розвиток» не є змістовними для аналізу самі по собі. Вони набувають значення в поєднанні з іншими словами (наприклад, «розвиток машинобудівної галузі» може виступати одним поняттям, вираженим словосполученням).

Виявивши наявність зв'язку, наступною проблемою є дослідження відносної ваги зв'язків між категоріями. Значення ваги, що прямує до максимального значення 100, вказує на те, що переважна більшість інформації в тексті чи сукупності текстів, що стосується однієї теми, одночасно відноситься й до іншої. Невелика вага відображає факт слабого зв'язку між темами або ж узагалі його відсутності. Зв'язок між темами завжди двосторонній, проте значення ваги зв'язку однієї теми до іншої може відрізнятись від значення зв'язку тієї іншої до цієї першої. Подібна розбіжність може вказувати на те, що одна тема є підтемою іншої.

Сфери використання

О. Єрмаков та В. Плешко наводять такі можливі сфери застосування методу побудови семантичних мереж [16]:

1) Дослідження тематичного складу цільової колекції документів (наприклад, потоку новин за обраний період часу), виявлення ключових тем та їхніх сполучень (семантичних полів), а також моніторинг динаміки інформаційного потоку в часі з огляду на потужність висвітлення ключових тем та їхніх сполучень.

2) Пошук нової, неочікуваної інформації, що пов'язана з досліджуваним об'єктом чи темою. Цікаві відкриття може дати список зв'язків, які є непомітними при прочитанні. Асоціативні ланцюги, що поєднують теми дослідження, також є цікавими для аналізу.

3) Виявлення в документах підтверджень відомих і невідомих зв'язків між об'єктами (темами). Пошук конкретних документів, що розкривають певні зв'язки, дозволяє одержати детальне уявлення про характер відносин між об'єктами.

Процедура семантичного аналізу представляє результати в наочній формі, причому сучасні програми машинного аналізу дозволяють варіювати кількість тем, що відображаються, масштаб, кольорове оформлення тощо.

Машинна обробка текстів дозволяє знаходити та відображати фрагменти мереж, які цікавлять дослідника. При цьому пошук відбувається не тільки за заданим словом, а й з урахуванням зв'язків між темами, тому можливі запити на

побудову найкоротшого асоціативного ланцюжка між двома темами.

Одним із різновидів семантичної мережі є тематичне дерево, що дозволяє не тільки виявляти найбільш значимі поняття та зв'язки між ними, а й ієрархію цих понять, їхню значимість у тексті, ієрархію зв'язків між ними.

Сильні сторони

Побудова семантичних мереж дає можливість:

- починати аналіз, не висуваючи гіпотез і строго не задаючи методики аналізу. Через цю особливість аналіз семантичних мереж може виступати як процедура, що передують дослідженню і слугує для вироблення гіпотез;
- динамічно змінювати модель предметної області.

Слабкі сторони та оцінка методу

Для побудови адекватної моделі необхідна значна вибірка. Знаходження зв'язків, непомітних під час прочитання, краще спрацьовує на великих обсягах інформації. Проте метод можна використати й для одиничного документа. Утім він не буде надто інформативним, бо відтворюватиме логічну побудову тексту, очевидну й після простого прочитання.

Варто зауважити, що семантична мережа будується без урахування логіки побудови тексту, - така мережа відображає лише асоціативні зв'язки, що можна виділити в тексті. Образно кажучи, побудова семантичної мережі є аналогією роботи правої півкулі головного мозку, яка відповідає за образне та асоціативне мислення. При цьому ліва півкуля, що відповідає за логічне мислення, повністю ігнорується. Через такий підхід до аналізу можливі певні неточності в передачі змісту. Утім численні спроби побудувати аналізатор тексту, що ґрунтувався б на виявленні логічної структури, поки що не дали бажаного результату [17].

Реферування тексту

Логіка та процедура

Аналіз, що починається з підрахунку частот, може закінчуватися реферуванням тексту. При цьому з тексту вибирається множина фраз, що містять найвагомші поняття та найсильніші зв'язки. Саме вони несуть максимальну інформацію про текст. Фрази виводяться в тому порядку, у якому вони з'являються в тексті. Звичайно, подібний реферат потребує певного шліфування, проте вже дані фрази легше підігнати одна до одної, аніж шукати їх у документі вручну.

Можлива варіація цього методу аналізу, що передбачає відбір не фраз, а готових речень, у яких зустрічається найбільша кількість значимих елементів. Іноді ця процедура враховує зв'язок між поняттями [18].

Сфери застосування та слабкі сторони

Основною слабкою стороною цього методу є необхідність використання аналізаторів, що спираються на семантичну будову речень і текстів. Через складність будови деяких речень, їхній зміст може бути неправильно переданий у разі машинної обробки.

Проте для більшості граматичних конструкцій реферат дійсно відображає зміст документа. Цей метод особливо цінний для аналізу великих масивів інформації, коли досліднику, наприклад, необхідно прослідкувати зміну домінуючих тем у засобах масової інформації.

Оцінка методу

Нааявні зараз на ринку програми є недостатньо завершеними, хоча й не новими. Вони часто потребують спеціаліста, який має досвід роботи саме з цим програмним забезпеченням, уміє налаштовувати й може їх адекватно змінювати, трактувати результати дослідження, змінювати обсяг шаблонної вибірки тощо. І варто зауважити, що зі збільшенням складності процедури аналізу й операцій, залишених машині, збільшується ризик одержати неадекватні дані і прийти до хибних висновків. Для якісно проведеного контент-аналізу мало тільки техніки, потрібна методологія й досвід проведення подібних досліджень.

Створений як спосіб аналізу документів за допомогою підрахунку тих чи інших елементів, контент-аналіз наразі поєднує в собі велику кількість методів дослідження тексту, що включають як кількісні, так і якісні елементи дослідження.

При цьому думки дослідників стосовно сутності даного методу варіюються від суто якісного розуміння, запропонованого Розовим, до виключно кількісного визначення Берельсона. Проте більшість науковців схиляються до думки, що контент-аналіз поєднує риси якісного та кількісного методів дослідження. Якісні харак-

теристики методу розуміються неоднозначно різними дослідниками. Одні вбачають ознаки якісного дослідження в можливості аналізу нечисленних або ж навіть відсутніх у тексті категорій, інші – у свободі дослідника формулювати рамки дослідження, ще інші – у можливості проводити аналіз без попередньо сформульованих гіпотез.

Контент-аналіз на даному етапі розвитку електронно-обчислювальної техніки виконується із залученням спеціального програмного забезпечення, яке дозволяє проводити різноманітні операції з документом чи сукупністю документів для виявлення прихованого змісту. Подібні техніки пройшли достатньо довгий шлях еволюції й зараз продовжують розвиватися. Немає жодної методики, яка гарантовано давала б нетривіальний надійний результат аналізу документів. Для використання того чи іншого методу варто звернути увагу на логіку проведення аналізу й обмеження, притаманні цій техніці.

Машинний контент-аналіз дозволяє оперувати великими обсягами текстів. Основною слабкою стороною, що притаманна майже всім «просунутим» технікам контент-аналізу, є необхідність мати певну базу: різноманітні тезауруси, лексикони, семантичні схеми побудови речення та тексту. Якщо ж програма сама здатна «вчитися», то найкращі результати будуть отримані тільки після того, як вона зробить адекватну шаблонну вибірку.

Очевидним позитивом всіх статистичних методів обробки тексту є швидкість, із якою можна одержати результати стосовно величезної вибірки документів.

Окрім зазначених, техніки контент-аналітичного дослідження, звичайно, мають цілу низку як сильних сторін, так і обмежень у використанні, що притаманні вже кожному конкретному інструменту дослідження.

Тож контент-аналіз є самостійним методом збору адекватної інформації про соціальну реальність. Велика кількість технік і підходів до проведення аналізу свідчить як про популярність методу, так і про невизначеність його меж. А це, у свою чергу, є доказом доцільності більш уважного погляду на метод контент-аналізу та його техніки.

1. Костенко Н., Іванов В. Досвід контент-аналізу: Моделі та практики. – К.: Центр вільної преси, 2003.
2. Ядов В. А. Стратегия социологического исследования. – М., 1995.
3. Roberts C. Text analysis for the social sciences: methods for drawing statistical inferences from texts and

- transcripts. – Mahwah NJ: Lawrence Erlbaum Associates, 1997.
4. Розов Н. С. Методика контент-анализа и визуализации понимания философских текстов. [Доступ до документа 16 лютого 2007р.] через <<http://www.lib-main/met-hist.htm>>.

5. Roberts C. Other than counting words: a linguistic approach to content analysis // Social Forces.— 1989.— Vol. 68.— P. 162–178.
6. Паршин П. Контент-анализ // Энциклопедия «Кругосвет». [Доступ до документа 17 лютого 2007р. через <<http://www.krugosvet.ru/articles/92/1009214/1009214a1.htm>>].
7. Шалак В. И. Контент-анализ. История метода // Проект ВААЛ online, 2004. [Доступ до документа 16 лютого 2007р. через <<http://www.vaal.ru/show.php?id=80>>].
8. Шалак В. И. Контент-анализ и теория ошибочных действий // Проект ВААЛ online, 2004. [Доступ до документа 16 лютого 2007р. через <<http://www.vaal.ru/show.php?id=84>>].
9. Singletary L. Empirical study of attributes and perceived benefits of applications integration for enterprise systems: Ph.D. dissertation.— New Jersey University, 2003.
10. Горбичишина С. Кое-что о методах социологических исследований // Аналитика.— 2002.— № 4.— С. 7–11.
11. Воронов Ю. П. Чтение между строк (контент-анализ в конкурентной разведке и не только в ней) // ЭКО.— 2005.— №14.— С. 50–64.
12. Федотова Л. Н. Анализ содержания – социологический метод изучения средств массовой коммуникации.— М.: Научный мир, 2000.
13. Григорьев С. И. Проведение контент-анализа // Основы современной социологии.— Алтайский гос. университет, 2001.
14. Манаев О. Т. Контент-анализ – описание метода // Библиотека Psyfactor, 2002. [Доступ до документа 16 лютого 2007р. через <<http://psyfactor.org/lib/kontent.htm>>].
15. Толстова Ю. Н. Измерение в социологии: Курс лекций.— М.: ИНФРА-М, 1998.
16. Ермаков А. Е., Плешко В. В. Семантическая сеть текста в задачах аналитика // Информатизация и информационная безопасность правоохранительных органов: XI Международная научная конференция: Сборник трудов.— М., 2002.
17. West M. Theory, method and practice in computer content analysis.— Westport, CT: Ablex. Place of Publication, 2001.
18. Ермаков А. Е., Плешко В. В. Синтаксический разбор в системах статистического анализа текста // Информационные технологии.— 2002.— №7.— С. 100–115.

V. Bondar, M. Dopira

REVIEWING OF METHOD OF CONTENT-ANALYSIS IN THE CONTEXT OF QUANTITATIVE AND QUALITATIVE METHODS OF REALIZATION

An endeavor to analyze the content analysis as a quantitative and qualitative method in sociological investigation was made in this article. The most widespread methods, logic of their analysis, scopes of application and possibilities were reviewed.