

BARBARA CYREK

UNIwersytet Jagielloński
WYDZIAŁ ZARZĄDZANIA I KOMUNIKACJI SPOŁECZNEJ
E-MAIL: CYREK.BARBARA@GMAIL.COM

DATA ZGŁOSZENIA: 5.01.2018
DATA PRZYJĘCIA DO DRUKU: 30.04.2018

Problem fluktuacji danych w badaniach prowadzonych w cyberprzestrzeni – casus komentarzy internetowych

STRESZCZENIE

Artykuł podejmuje zagadnienie metodologii badań internetowych w kulturze dystrakcji i informacyjnej nadpodaży. Współczesna socjologia Internetu mierzy się nie tylko z niekontrolowanym przyrostem danych, ale również z ich zmiennością, która zachodzi w czasie rzeczywistym. W pracy omówiono zjawisko badania komentarzy internetowych, w przypadku których analiza fluktuacji danych daje informacje na temat mechanizmów komunikowania w cyberprzestrzeni. Jednocześnie podniesione zostały etyczne dylematy wykorzystania tego materiału badawczego.

SŁOWA KLUCZOWE

big data, netnografia, dane zmienne, dane stające się, badania Internetu

Słowo wstępne

Kłęska urodzaju danych wydaje się znamieną dla kultury cyfrowej i społeczeństwa informacyjnego. W globalnym kontekście społecznym hipertrofia informacji implikuje atrofię wrażliwości¹, jest katalizatorem zjawisk takich

¹ M. Szpunar, *Wrażliwość (nie tylko) artystyczna*, „Kultura i Społeczeństwo” 2017, nr 1, s. 123–134.

jak anestetyka (zniesienie zdolności doznawania)² i desensytyzacja (odwrażliwienie na bodźce)³. Nadmiar informacyjny powoduje także niedostatek uwagi⁴, która jest zasobem słabo powiększalnym⁵. Jednak jest ona kluczowa dla szeroko rozumianej ekonomii mediów, w przypadku których zainteresowanie odbiorców generuje zyski. Zagadnienie to porusza między innymi teoria agenda-setting Maxwella McCombsa i Donalda Shawa, w której agenda jest hierarchią priorytetów⁶. Tym samym uwaga staje się kategorią niezwykle pożądaną, szczególnie zważywszy na fakt, że stanowi ona filtr kształtujący w jednostce obraz rzeczywistości. Marek Krajewski określa dzisiejszą kulturę jako „kulturę dystrakcji” i stwierdza, że „jednostki nigdy nie działają na podstawie tego, jakie jest środowisko ich życia, ale raczej na podstawie tego, co w jego obrębie dostrzegają, na czym się koncentrują, co czynią istotnym. Uwaga stwarza dla jednostki rzeczywistość, bo decyduje o tym, co istnieje i jest obecne, ale też chroni ją przed nadmiarem informacji, a tym samym pozwala jej w ogóle działać”⁷. Jednak ta swoista ochrona przed przebodźcowaniem nie oznacza nieświadomości istnienia treści, które nie zostały zaabsorbowane. W praktyce może to oznaczać, że świadome doświadczanie ma potencjał przekraczania zbioru tych informacji, na których zogniskowana jest uwaga⁸. Jak trafnie stwierdzają Michał Wierzchoń i Aleksandra Gruszka, „przyglądając się jakiejś złożonej scenie, nasza uwaga koncentruje się jedynie na niewielkim fragmencie doświadczenia percepcyjnego. Niewątpliwie jednak mamy wrażenie dostępu do szeregu innych informacji, na których uwaga nie została skoncentrowana. A zatem do pewnego stopnia przetwarzamy również świadomie informacje, na których nie koncentruje się nasza uwaga”⁹.

² W. Welsch, *Estetyka i anestetyka*, tłum. M. Łukasiewicz, [w:] *Postmodernizm. Antologia przekładów*, red. R. Nycz, Kraków 1998, s. 520–546.

³ M. Szpunar, *Wrażliwość...*, op. cit.

⁴ H. Simon, *Designing Organizations for an Information-Rich World*, [in:] *Computers, Communications, and the Public Interest*, ed. M. Greenberger, Baltimore 1971, s. 40–41.

⁵ M. Szpunar, *Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych a reprodukowaniem schematów komunikowania masowego*, Warszawa 2012, s. 125–133.

⁶ M. E. McCombs, D. L. Shaw, D. H. Weaver, *New Directions in Agenda-setting Theory and Research*, „Mass Communication and Society” 2014, Vol. 17, No. 6, s. 781–802.

⁷ M. Krajewski, *Kultura dystrakcji – deficyty uwagi i strategie jej kumulacji*, [online] http://krajewskimarek.blox.pl/resource/Ekonomia_uwagi.doc [dostęp: 3.01.2017].

⁸ V. A. Lamme, *Why Visual Attention and Awareness Are Different*, „Trends in Cognitive Sciences” 2003, No. 7 (1), s. 12–18; H. C. Lau, D. Rosenthal, *Empirical Support for Higher-order Theories of Conscious Awareness*, „Trends in Cognitive Sciences” 2011, No. 15 (8), s. 365–373.

⁹ M. Wierzchoń, A. Gruszka, *Czy uwaga ekstensywna jest świadoma? O relacjach uwagi i świadomości w kontekście teorii stanów uwagi*, „Studia Psychologiczne” 2011, t. 49, z. 3, s. 45.

Nadpodaż danych a badania internetowe

Świadomość faktu, że coś zostało pominięte, że jakimś treściami czy zagadnieniom nie poświęcono dostatecznie dużo uwagi, jest szczególnie ważna w nauce. Naukowiec świadomy swoich ograniczeń zachowuje w konstruowaniu wniosków większą ostrożność, a warto pamiętać, że pochopne dokonywanie uogólnień jest bliższe wiedzy potocznej niż naukowej¹⁰. Co więcej, sam dobór metodologii badawczej wiąże się z pytaniem o neutralność aksjologiczną badacza¹¹. Zagadnienie to jest niezwykle istotne w świetle występowania w mediach cyfrowych nadpodaży danych, których obfitość (przy korzystaniu z klasycznych metod badawczych) wiąże się z koniecznością ich ograniczenia. Z pomocą przychodzą tu mechanizmy analizy *big data*, czyli danych, które – jak pisze Maria Mach-Król – „pochodzą ze źródeł nieustrukturyzowanych, takich jak strumień kliknięć w www, sieci społecznościowe (Twitter, blogi, Facebook), nagrania wideo ze sklepów, nagrania rozmów w call center itp., informacje w czasie rzeczywistym z czujników [...] i innych urządzeń identyfikujących”¹². Magdalena Wójcik podkreśla, że terminem *big data* określa się duże, nieuporządkowane i dynamicznie zmieniające się zbiory danych¹³. Analiza skupień, modele predykcyjne czy analiza reguł asocjacyjnych to tylko niektóre tradycyjne techniki *data mining*¹⁴. Mechanizmy wydobywania informacji z gąszczu danych nie wymagają ich ograniczenia, co więcej – pierwsze definicje *big data* mówią o maksymalizowaniu liczby danych celem wydobycia wartości informacyjnych¹⁵.

Na drugim biegunie badań prowadzonych w Internecie lokuje się etnografia internetowa, tak zwana netnografia. Badania internetowe początkowo służyły triangulacji danych, uzupełnieniu tradycyjnej antropologii czy etnografii analizą zjawisk zachodzących w cyberprzestrzeni rozumianej jako przedłużenie rzeczywistości pozawirtualnej. Niemniej, jak słusznie uważa Dariusz Jemielniak,

¹⁰ M. Szpunar, *Nauka przeciw fałszerstwom?*, [w:] *Człowiek – media – edukacja*, red. J. Morbitzer, E. Musiał, Kraków 2013, s. 380–387.

¹¹ Ibidem.

¹² M. Mach-Król, *Analiza i strategia big data w organizacjach*, „Studies & Proceedings of Polish Association for Knowledge Management” 2015, Vol. 74, s. 41.

¹³ M. Wójcik, *Big data w zarządzaniu informacją – przegląd wybranych zagadnień*, [w:] *Inspiracje i innowacje: zarządzanie informacją w perspektywie bibliologii i informatologii*, red. S. Cisek, Kraków 2016, s. 64.

¹⁴ M. Mach-Król, op. cit., s. 51.

¹⁵ M. Tabakow, J. Korczak, B. Franczyk, *Big Data – definicje, wyzwania i technologie informatyczne*, „Informatyka ekonomiczna” 2014, nr 1 (31), s. 141.

[...] badanie zachowań wyłącznie wirtualnych (bez prób łączenia ich z zachowaniami tych samych osób z dala od komputera, a nawet bez badania zachowań tych samych osób w różnych społecznościach wirtualnych) ma o tyle praktyczny sens, że zachowania wirtualne niekoniecznie zawsze mają związek z trwałymi zachowaniami i postawami offline. Oznacza to, że etnografie wirtualne [...] otwierają dla antropologów i badaczy organizacji nowe pole badawcze i możliwości analizy zachowań, które wcześniej nie występowały lub nie dawały się badać¹⁶.

Zarówno etnografia tradycyjna, jak i netnografia jako podstawową technikę wykorzystują obserwację: czynną, opartą na uczestnictwie biernym lub peryferyjnym albo też wariant pośredni¹⁷. Badania internetowe w naukach społecznych mają charakter dwojaki: są to badania medium, jakim jest sieć, lub też badania prowadzone drogą elektroniczną¹⁸. Choć nie jest to oczywiste, obie formy badań możliwe są do przeprowadzenia dzięki internetowym komentarzom, które badacz nie tylko analizuje, ale również może za ich pomocą wywoływać dane, jednocześnie dokonując wirtualnej obserwacji następstw swoich działań.

Analiza komentarzy może być stosowana z wykorzystaniem algorytmów dedykowanych *big data*, jest też często wykorzystywana w badaniu medialnych dyskursów. W drugim wypadku nadpodaż danych nie jest zjawiskiem pożądanym. Jak zwróciła uwagę Małgorzata Lisowska-Magdziarz, w analizie dyskursu „znacznie istotniejsza jest dokładność, wszechstronność i szczegółowość badań niż duży rozmiar analizowanego materiału. Lepiej zatem wybrać materiał mniejszy, lecz możliwy do dokładnego zbadania”¹⁹.

Co jednak znamienne dla internetowych komentarzy, ich przyrost nie jest kontrolowany. Nie sposób ocenić, kiedy przestanie ich przybywać (lub ubywać). Warto przy tym pamiętać, że badania przekazu bazują na analizie treści, która, jak określił to Bernard Berelson, jest techniką badawczą służącą „obiektywnemu, systematycznemu i ilościowemu opisowi jawnej zawartości komunikatów”²⁰. Definicja ta budzi pewne wątpliwości, szczególnie dotyczące obiektywizmu oraz ilościowego charakteru. Zwraca na to uwagę Karolina Szczepaniak, która słusznie stwierdza, iż obiektywizm może tu być

¹⁶ D. Jemielniak, *Netnografia, czyli etnografia wirtualna – nowa forma badań etnograficznych*, „Prakseologia” 2013, nr 154, s. 98–99.

¹⁷ K. Jurek, *Badania społeczne w internecie. Wirtualna etnografia w teorii i praktyce*, „Nauka i Szkolnictwo Wyższe” 2013, nr 1 (41), s. 90.

¹⁸ M. Szpunar, *Badania Internetu vs. badania w Internecie, czyli jak badać nowe medium – podstawowe problemy metodologiczne*, „Studia Medioznawcze” 2007, nr 2 (29), s. 80–89.

¹⁹ M. Lisowska-Magdziarz, *Analiza tekstu w dyskursie medialnym: Przewodnik dla studentów*, Kraków 2007, s. 44.

²⁰ Cyt. za: ibidem, s. 13.

jedynie postulatem. Konieczne jest „zachowanie świadomości, że zewnętrzne warunki społeczno-kulturowe, wewnętrzne przekonania i hierarchia wartości, a nawet sam cel badań i motywacje do ich podjęcia (każdorazowo zindywidualizowane) mogą mieć, i zazwyczaj mają, wpływ na samego badacza i wyniki jego analiz”²¹. Z kolei – jak konstatuje badaczka – „sama «ilościowość» opisu może okazać się w niektórych kwestiach nie tylko niewystarczająca, ale i zawodna”²². Analiza treści pozwala na ocenę „zawartości przekazów informacyjnych z punktu widzenia pewnych standardów przyjętych przez badacza”²³. Standardy te muszą być przejrzyste, a metoda możliwa do powtórzenia. Zarówno proces badawczy, jak i analityczny muszą – jak zawsze w badaniach społecznych – cechować się transparentnością²⁴. Stąd analiza treści, będąca podstawą badań przekazów medialnych, sama w sobie jest techniką niełatwą i wymagającą konsekwencji oraz skrupulatności. Fluktuacja analizowanego materiału badawczego nie czyni zadania łatwiejszym.

Struktura komentarzy często nastrocza badaczowi trudności. Komentarze do tekstu (lub jego fragmentów) mieszają się z komentarzami do innych komentarzy, stąd też prześledzenie toku dyskusji może być czasochłonne, a w przypadku utraty części danych nawet niemożliwe. W świetle teorii sieci społecznych, czyli struktur relacji pomiędzy społecznymi podmiotami²⁵, należy zwrócić uwagę również na wzajemne powiązanie internetowych treści, w których funkcjonowaniu szczególną rolę mogą odgrywać nie tyle znaczenia, ile przepływy. Uwikłanie w pewną strukturę społeczną warunkuje współzależność pomiędzy podmiotami, tym samym sieć jest niezwykle czułą, „żywą tkanką”, odpowiadającą na wszelkie, nawet nieznaczne zmiany, które w jej obrębie dokonują się na różnych poziomach²⁶. Tym samym zmiana w komentowanym tekście lub zmiana w jednym komentarzu może spowodować kaskadę edycji i usunięć innych komentarzy, zatem przyczynić się do całkowitej inwersji w materiale badawczym.

²¹ K. Szczepaniak, *Zastosowanie analizy treści w badaniach artykułów prasowych – refleksje metodologiczne*, „Acta Universitas Lodzianensis: Folia Sociologica” 2012, t. 42, s. 85.

²² Ibidem.

²³ D. P. Cartwright, *Zastosowania analizy treści*, [w:] *Metody badań socjologicznych*, oprac. S. Nowak, Warszawa 1965, s. 150.

²⁴ E. Babbie, *Badania społeczne w praktyce*, tłum. W. Betkiewicz et al., Warszawa 2006, s. 522.

²⁵ S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge 2009.

²⁶ N. M. Moolenaar, *A Social Network Perspective on Teacher Collaboration in Schools: Theory, Methodology and Applications*, „American Journal of Education” 2012, No. 1, s. 10–11.

Młodsze pokolenia internautów – cyfrowi tubylcy – ma trudności z pisaniem i czytaniem długich, linearnych tekstów²⁷. Implikuje to sytuację, w której komentarze pisane są chaotycznie, bez interpunkcji, często z użyciem skrótów lub nawet mają postać wyłącznie linku. Jak pisze Szpunar: „zamiast argumentować, przesyłamy dyskutantowi linka, zwalniając się niejako z konieczności udowadniania jakiejś hipotezy, przerzucając tę odpowiedzialność na zewnętrzny zasób, traktowany dodatkowo przez niektórych jako obiektywny. Ciężar dowodowy spoczywa zatem nie na nas, a na wyszukiwarkach”²⁸. Komunikacja za pomocą linków nastrocza trudności badaczom Internetu, również w przypadku komentarzy – czy wartość merytoryczną takiej wypowiedzi należy ocenić na podstawie wartości merytorycznej komunikatu, do którego ona prowadzi? Być może warto tak skonstruowane komentarze klasyfikować osobno.

Ponadto, jak konstatuje William Mitchell, „nowoczesność pociąga za sobą hegemonię widzenia i mediów wizualnych”²⁹. Wzrokocentryzm skutkuje kondensacją warstwy słownej, a często komentarze mają postać wyłącznie grafiki. Zjawisko to, szczególnie przy stosowaniu analizy treści, wymaga dekodowania znaczeń w opublikowanych obrazach, a tym samym sięgnięcia po metodologiczny dorobek badań wizualnych.

Wybór komentarzy jako źródła danych rodzi liczne zapytania metodologiczne, a rozmaite podejścia badawcze kładą nacisk na inne aspekty analizy. W przypadku *big data* szczególne znaczenie będzie mieć ilość – klasyczne definicje zakładają tutaj, że im więcej danych do zbadania, tym lepiej. Na przeciwnym biegunie lokuje się analiza dyskursu, w perspektywie której nadmiar materiału nie jest zjawiskiem pożądanym – ważniejsze staje się jego wielopłaszczyznowe, wszechstronne i wnikliwe przebadanie. Krytyczna analiza dyskursu dąży do badania ukrytych relacji przyczynowości pomiędzy praktykami dyskursywnymi a szerszymi strukturami społecznymi i kulturowymi. Z kolei analiza treści bazuje na jawnej zawartości komunikatów – kontekst nie musi być tutaj znany. Nie istnieje jedyna słuszna metodologia analizy komentarzy internetowych. Dobór metody warunkowany jest przez hipotezy badawcze i to badacz sam musi określić, jakich efektów swojej pracy oczekuje.

²⁷ M. Prensky, *Digital Natives, Digital Immigrants*, [online] <https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf> [dostęp: 1.12.2017].

²⁸ M. Szpunar, *Imperializm kulturowy Internetu*, Kraków 2017, s. 86.

²⁹ W. J. T. Mitchell, *Pokazując widzenie: krytyka kultury wizualnej*, „Artium Quaestiones” 2006, nr XVII, s. 280.

Komentarze internetowe stanowią specyficzne źródło danych ze względu na ich różnorodność, a często również zmienność i nietrwałość. Wstępne zapoznanie się z materiałem badawczym może pozwolić na określenie różnorodności komentarzy w ich warstwie formalnej. To wstępne ustalenie, czy w komentarzach prócz linearnego tekstu występują linki, obrazy, gify, emotikony lub hasztagi, pozwoli skonstruować narzędzia do analizy niejednolitego materiału. Jednak bez względu na formę badacz musi liczyć się z fluktuacją danych, które niemal zawsze mogą zostać utracone lub zmienione.

Fluktuacja danych – casus komentarzy

Analiza internetowych komentarzy jest przedsięwzięciem na pozór łatwym, nie wymaga bowiem specjalistycznych narzędzi. Może ją przeprowadzić każdy, kto posiada dostęp do sieci. Materiał badawczy jest niezwykle cenny, gdyż stanowią go wypowiedzi występujące w sytuacjach naturalnych, zatem – jak pisze Małgorzata Lisowska-Magdziarz – „prawdziwe, niespreparowane, niepokawałkowane, niedopasowane uprzednio do celu naszych badań”³⁰. Szczególnie istotny jest tu klucz kategoryzacji danych, który należy przyjąć w zależności od celu przeprowadzanych badań, a tym samym techniki badawczej (analiza językowa, analiza frekwencyjna etc.). Słabą stroną takich badań jest niemożność scharakteryzowania komentujących pod względem społeczno-demograficznym³¹, a często nawet ocena, czy pod kilkoma różnymi nickami³² nie kryje się ta sama osoba – ponieważ cyberprzestrzeń w szerokim zakresie funduje anonimowość. Zadanie to ułatwić może połączenie licznych stron www na przykład z Facebookiem czy kontem Google+, dzięki któremu komentujący zamiast nickiem podpisany jest imieniem i nazwiskiem. Ułatwia to dotarcie do autora komentarza. Niemniej podstawowe umiejętności informatyczne, takie jak na przykład określenie adresu IP nadawcy, wydają się kluczowe w badaniu komentarzy internetowych. Co więcej, bez wyspecjalizowanych narzędzi nie sposób nawet określić, czy autorem komentarza jest człowiek, czy też stworzony w tym celu program, dlatego też badacz powinien mieć świadomość słabych stron przeprowadzanej przez siebie analizy i wszelkie zastrzeżenia zawrzeć pod-

³⁰ M. Lisowska-Magdziarz, *Analiza tekstu...*, op. cit., s. 25.

³¹ D. Jemielniak, op. cit.

³² Nick (*nickname*) to określenie pseudonimu, loginu, identyfikatora lub nazwy użytkownika na portalach, forach internetowych, a także podczas logowania do rozmaitych usług, np. poczty e-mail.

czas opracowywania wniosków. Można się zatem zgodzić, iż zasadniczą trudnością w przeprowadzeniu rzetelnych i obiektywnych badań nad komentarzami internetowymi jest dobór próby. Ważne jest nie tylko źródło komentarzy, ale również czas i sposób ich gromadzenia, a także ich specyfika (tekst, link, emotikon, grafika lub połączenie kilku form).

Casus 1: dane wywoływane³³ przez badacza

Jeżeli wywoływanie danych następuje w internetowych serwisach do tego przeznaczonych, takich jak na przykład kreatory ankiet, wówczas badacz sam decyduje o terminie zamknięcia systemu i zakończeniu badania. W przypadku komentarzy szczególnie cenna jest naturalność sytuacji wyrażenia opinii przez badanych. Tym samym ustrukturyzowane serwisy służące zbieraniu danych nie mogą zostać wykorzystane. Zamiast tego w roli przestrzeni badawczej sprawdzą się blogi, serwisy informacyjne czy media społecznościowe. Te ostatnie wypracowały własne, prymitywne narzędzia do zbierania danych – coraz częściej stanowią one nie tylko pole, ale i instrumentarium badawcze³⁴. W tych wirtualnych przestrzeniach istnieją dwie podstawowe możliwości wywołania komentarzy:

- a) stworzenie własnego tekstu³⁵ i opublikowanie go w sieci z możliwością komentowania,
- b) napisanie komentarza, który sprowokuje innych internautów do wypowiedzi.

Taki sposób wywoływania danych pozwala na uzyskanie komentarzy w interesującym badacza obszarze. Posunięcie to wydaje się szczególnie uzasadnione w przypadku badania zagadnień o niewielkiej popularności. Należy jednak bardzo ostrożnie podejść do wyboru miejsca publikacji tek-

³³ W nauce mówi się o danych wywołanych i zastanych. Ponieważ przyrost komentarzy jest niekontrolowany i może trwać (poza pewnymi szczególnymi przypadkami opisanymi w dalszej części artykułu) w zasadzie bez końca, w tym przypadku proponuję poślubiwać się formą niedokonaną.

³⁴ B. Cyrek, *Media społecznościowe – nowa przestrzeń nauki*, „Kognitywistyka i Media w Edukacji” 2016, nr 2, s. 45–56.

³⁵ Może to być również tekst kogoś innego, opublikowany przez badacza w swoistym kontekście, specyficznym czasie, opatrzony nietypowym komentarzem etc. Tekst rozumiany jest tutaj jako każde zdarzenie komunikacyjne: tekst pisany, film, obraz etc. – zob. M. Rygielska, *O „tekście kultury”*, „Zeszyty Etnologii Wrocławskiej” 2015, nr 1, s. 27–43; P. Garncarek, *Teksty kultury w nauczaniu języka polskiego jako obcego*, „Acta Universitatis Lodziensis: Kształcenie Polonistyczne Cudzoziemców” 2010, nr 17, s. 99.

stu, ponieważ może ono determinować wydzźwięk pojawiających się komentarzy.

Wywoływanie komentarzy wzbudzać może wątpliwości natury etycznej. Czy zachęcanie internautów do komentowania bez informowania ich o tym, że ich wypowiedzi stanowią materiał badawczy, jest z etycznego punktu widzenia dopuszczalne? Nie bez znaczenia jest tu fakt, że komentujący wypowiadają się publicznie, zatem świadomi są tego, że ich słowa może czytać i wykorzystać każdy inny internauta. Kolejne zastrzeżenia wynikają z pytania, czy wypowiedzi powstałe w odpowiedzi na komentarz badacza mają taką samą wartość jak te, które pojawiły się niesprowokowane. Naukowiec musi mieć świadomość, że dane te bez jego ingerencji nigdy by nie powstały. Mamy wprawdzie do czynienia ze stosunkowo nowym, nieznanym kłasykiem socjologii materiałem badawczym, niemniej jednak tego typu problemy etyczne są doskonale znane choćby z obserwacji uczestniczącej niejawniej, której wiele uwagi poświęca Kodeks Etyki Socjologa. Wśród zawartych tam zasad znajduje się następująca informacja: „[...] w niektórych przypadkach zastosowanie tej formy badań może być uzasadnione. Na przykład wówczas, gdy można przypuszczać, iż uczestnicy, wiedząc, że są badani, zmienią swoje zachowanie”³⁶. Zagadnienie to podejmują również Martyn Hammersley i Paul Atkinson, zdaniem których rzeczywistość społeczną należy badać w jej naturalnym, niezakłóconym przez badacza stanie³⁷. Jawność badania może bowiem wpłynąć na jego wynik. Tak zwanego efektu Hawthorne³⁸ można uniknąć, informując badanych dopiero po przeprowadzeniu badania, celem uzyskania zgody na wykorzystanie ich wypowiedzi.

Fluktuacja danych zawartych w komentarzach wywoływanych przez badacza może się pojawiać na kilku poziomach, co jest charakterystyczne dla wszelkich nieustrukturyzowanych internetowych narzędzi badawczych. Po pierwsze, zasadniczą kwestią jest moment zaprzestania pobierania materiału badawczego. Przyrost danych jest tu trudny do kontrolowania (chyba że badacz jest autorem komentowanego tekstu i w pewnym momencie usunie go lub wyłączy możliwość komentowania). Nie można kategorycznie stwierdzić, kiedy dane przestaną napływać lub kiedy ich napływ będzie następował w ilościach na tyle niewielkich (lub stosunkowo niewielkich), że zaprzestanie pobierania materia-

³⁶ *Kodeks etyki socjologa*, [online] <http://pts.org.pl/wp-content/uploads/2016/04/kodeks.pdf> [dostęp: 9.03.2018].

³⁷ M. Hammersley, P. Atkinson, *Metody badań terenowych*, tłum. S. Dymczyk, Poznań 2000, s. 16.

³⁸ S. R. G. Jones, *Was There a Hawthorne Effect?*, [online] <https://www.economics.mcmaster.ca/research/publications/working-papers/1985-1999/91-01.pdf> [dostęp: 7.03. 2018].

łu nie wpłynię znacząco na wynik badania. Wydaje się, że w kontekście kultury nadmiaru³⁹ i kryzysu uwagi można założyć, iż „cykl życia” tekstu, do którego komentarze bierze się na warsztat badawczy, jest współcześnie krótki. Jednak cyberprzestrzeń rządzi się prawami hipertekstu – nie sposób przewidzieć, czy nagłe zdarzenie (skandal, katastrofa, wypowiedź autorytetu) nie wznowi dyskusji nad danym problemem, a w sieci linków uwaga internautów nie zostanie ponownie skierowana na tekst, pod którym gromadziliśmy komentarze do analizy. W tym kontekście wartościowe wydaje się powrócenie do danego tekstu po jakimś czasie i ponowna weryfikacja komentarzy, co z pewnością niesie potencjał poznawczy. Jak podkreśla Magdalena Szpunar: „dzięki danym zebranych w różnym czasie możemy lepiej rozumieć kontekst historyczny, opisywać i wyjaśniać zmiany”⁴⁰.

Kolejną zasadniczą kwestią fluktuacji danych zawartych w komentarzach jest ich e d y c j a. Zjawisko to jest powszechne, ponieważ obecnie większość portali internetowych dopuszcza możliwość wprowadzenia zmian w opublikowanym przez siebie komentarzu. Rozwiązaniem tego problemu może być zapisywanie zrzutów ekranu i porównanie ewentualnych zmian w treści poszczególnych wypowiedzi. Nawet jeśli historia edycji jest widoczna i z łatwością możliwa do przesledzenia, badacz musi liczyć się z faktem, że zmiany zachodzą mogą w czasie rzeczywistym i nigdy nie ma całkowitej pewności co do tego, że wszystkie edycje zostały uwzględnione⁴¹. Przypadek ten również nastrocza wątpliwości natury etycznej: skoro internauta zmienił swą wypowiedź, czy należy uwzględniać jej pierwotną wersję? Warto także zastanowić się nad przyczyną dokonanej edycji – czy na przykład była to reakcja na wypowiedź innego internauty? Wydaje się, że rzetelnych informacji w tym zakresie nie sposób uzyskać bez skontaktowania się z autorem edytowanej wypowiedzi.

Istotną (jeśli nie najważniejszą) kwestią jest również zjawisko u t r a c e n i a d a n y c h, kiedy autor komentarza skasuje swoją wypowiedź⁴².

³⁹ T. Szlendak, *Kultura nadmiaru w czasach niedomiaru*, „Kultura Współczesna” 2013, nr 1, s. 7–26.

⁴⁰ M. Szpunar, *Internet w procesie gromadzenia danych o charakterze wtórnym*, „Studia Medioznawcze” 2009, nr 2 (37), s. 141.

⁴¹ Wyjątkiem jest zupełne zamknięcie napływania i edycji danych, tj. usunięcie tekstu, pod którym zbierano materiał badawczy, ewentualnie zablokowanie możliwości komentowania, co na wielu portalach równoznaczne jest z ukryciem wszystkich dotąd opublikowanych komentarzy.

⁴² W przypadku wywoływania danych za pomocą komentarza badacza pod publikacją innej osoby komentarze internautów mogą zostać usunięte przez administratora danej witryny. Przypadek ten zostanie opisany w dalszej części artykułu.

Sytuacji tej pozwalają uniknąć zrzuty ekranu. Niemniej po raz kolejny pojawia się tu dylemat etyczny: czy można wykorzystać w badaniu wypowiedź, z której internauta się wycofał? Jak zauważa Magdalena Szpunar, ochrona praw internautów biorących udział w badaniach prowadzonych w sieci powinna polegać między innymi na zapewnieniu możliwości wycofania się z badania na każdym etapie jego realizacji⁴³. Jednak w przypadku obserwacji ukrytej kwestia ta wydaje się znacznie bardziej skomplikowana. Sposobem rozwiązania tego etycznego konfliktu – zarówno w przypadku edycji, jak i usunięcia danych – może być uwzględnienie ilości komentarzy zmienionych/usuniętych w statystyce ilościowej, co pokaże rozmiar problemu, jednak bez włączenia do analizy jakościowej treści edytowanych bądź skasowanych przez autorów. Innym rozwiązaniem może być wspomniane już odtajnienie badania *post factum* i próba uzyskania zgody na wykorzystanie materiału uzyskanego w sposób niejawnny.

Casus 2: dane wywoływane przez inną osobę

W przypadku analizy komentarzy wywołanych przez inną osobę, tj. umieszczonych pod tekstem, który nie został przez badacza w tym celu zaprojektowany, niezwykle ważna jest świadomość specyfiki materiału badawczego.

Ważną rolę odgrywa tu sam autor publikacji⁴⁴, o której opinie brane są na warsztat badawczy. Po pierwsze, osoba taka może – niczym badacz wywołujący dane pod swoim tekstem – sama komentować swoją publikację celem wzbudzenia dyskusji. Co ważne, nie ma pewności, czy nie zrobi tego anonimowo. Jeżeli autor publikuje komentarze w sposób jawny, to znaczy bierze udział w dyskusji nad swoją publikacją, należy jego wypowiedzi potraktować oddzielnie od reszty komentarzy. Być może warto osobnej kategoryzacji poddać również opinie publikowane w odpowiedzi na komentarze autora – które być może nie pojawiłyby się, gdyby ten nie wziął udziału w dyskusji. Jeżeli jednak eksponent pozostanie anonimowy, zdemaskowanie go wydaje się niemożliwe. Zastrzeżenie to warto ująć w opisie prowadzonych badań, bowiem jego świadomość rzuca nowe światło na materiał badawczy.

Wspomniany wcześniej „cykl życia” tekstów w Internecie ma niebagatelne znaczenie w przypadku analizy komparatystycznej. Porównanie ko-

⁴³ M. Szpunar, *Badania Internetu...*, op. cit., s. 89.

⁴⁴ Wypada podkreślić, iż autor publikacji niekoniecznie musi posiadać prawa autorskie do tekstu, który stanowi jej treść. Może to być również wykorzystanie cudzego dzieła w rozmaity sposób.

mentarzy pod kilkoma publikacjami powinno uwzględniać datę upublicznienia ich w sieci. Jeżeli jeden tekst widnieje w Internecie od miesięcy, drugi natomiast zaledwie kilka dni, istnieje prawdopodobieństwo, że pod tym drugim pojawią się jeszcze nowe, cenne opinie. Kwestia tego, kiedy zaprzestać pobierania materiału badawczego, jest niezwykle trudna. Badacz sam musi dokonać oceny, uwzględniając aktualne wydarzenia z danej dziedziny lub na przykład średni czas znaczącego przyrostu ilości komentarzy na danym portalu (dzięki porównaniu czasu dodawania komentarzy pod innymi publikowanymi tam treściami).

Prócz niekontrolowanego przyrostu danych w tym wypadku badacz musi się liczyć także z problemem ich edycji bądź utracenia. W niektórych serwisach historia edycji komentarza widoczna jest wyłącznie dla administratora strony. W takich wypadkach zewnętrzny obserwator zmuszony jest bazować na zrzutach ekranu, jeżeli pragnie prześledzić zmiany w badanych wypowiedziach. Zagadnienie utracenia danych jest tutaj znacznie bardziej newralgiczne.

W przypadku brania na warsztat komentarzy wywołanych przez inną osobę szczególnie istotny jest fakt, że mogą one zostać usunięte nie tylko przez ich autorów, ale również przez szeroko rozumianych administratorów danej witryny: czy to w mediach społecznościowych, czy na blogach, czy w mediach informacyjnych etc. Pierwszym zagrożeniem jest tutaj usunięcie całej komentowanej publikacji lub nagłe wyłączenie możliwości jej komentowania – jeżeli badacz nie zapisywał danych na dysku, wówczas utraci je bezpowrotnie. Drugi wariant jest znacznie bardziej skomplikowany i wiąże się z licznymi mechanizmami komunikacji w cyberprzestrzeni.

Dobrym przykładem są tu komentarze zamieszczane na Facebooku, które narażone są na usunięcie z wielu stron: przez samych komentujących, przez administratorów (grup, wydarzeń, stron fanowskich), a także przez moderatorów Facebooka, czyli z pewnego metapoziomu. Przykład Facebooka nie jest odosobniony – podobne mechanizmy narażenia na utracenie danych zachodzą na rozmaitych portalach, nie tylko społecznościowych, ale również informacyjnych etc. Należy liczyć się z tym, iż wiele serwisów zawiera w regulaminach klauzulę o prawie do usuwania komentarzy wulgarynych, obrażających uczucia religijne etc. Pytanie, które musi sobie zadać sam badacz, brzmi następująco: Czy komentarze *ad personam*, zawierające inwektywy lub nieodnoszące się zupełnie do treści publikacji rzeczywiście wnoszą coś wartościowego do badań?

Nie jest oczywiście przesądzone, że usunięte wypowiedzi miały niską wartość merytoryczną. Należy pamiętać o regule homofilii, prowadzącej do

homogenizacji środowiska społecznego⁴⁵, której mechanizmy zdaniem Jamesa Fishkina wzmacniane są przez cyberprzestrzeń⁴⁶. W praktyce oznacza to, że moderator danej witryny może usunąć wypowiedź choćby dlatego, że nie przystaje ona do powszechnego na tej stronie dyskursu⁴⁷.

W przypadku usuwania własnych komentarzy można mówić o *oddolnej utracie danych*, natomiast ich kasowanie przez administratorów będzie utratą *odgórną*. Wydaje się, iż ta druga sytuacja nie niesie z sobą bagażu etycznych dylematów: komentujący nie wycofał się ze swojej opinii, lecz osoba trzecia usunęła jego głos z toczącej się dyskusji. Usunięte wypowiedzi mogą stanowić niemały odsetek całości toczącej się w komentarzach polemiki. Jest to osobne zagadnienie, które powinno wzbudzić zainteresowanie badacza. Odpowiedzi na pytania „ile komentarzy ilu osób usunięto?” oraz „co było w ich treści?” pozwolą poznać panujące na danej stronie mechanizmy zarządzania dyskusją. Dotarcie do takich danych nie jest zadaniem łatwym, ale nie jest też niemożliwe. Pomagają tutaj zrzuty ekranu, jak również informacje pochodzące z samego serwisu, na przykład komunikat o tym, że „komentarz został usunięty przez...” lub – stosowane na przykład na Facebooku – oznaczenia graficzne⁴⁸. Niewątpliwą wartością badania będą dane na temat fluktuacji danych w okresie ich zbierania – nie tylko ilościowe (zmieniająca się liczba komentarzy), ale również jakościowe (zmieniające się treści). Badacz może – a w niektórych sytuacjach nawet powinien – próbować poznać motywy usuwania wypowiedzi (własnych lub cudzych), kontaktując się z osobami, które tego dokonały. Wiedza taka ma ogromny potencjał poznawczy w naukach o komunikacji społecznej, w szczególności o mechanizmach komunikowania drogą online.

⁴⁵ M. Szpunar, *Imperializm...*, op. cit., s. 34.

⁴⁶ *Debating Deliberative Democracy*, eds. J. Fishkin, P. Laslett, Malden 2006.

⁴⁷ Zjawisko to przedstawiłam w grudniu 2016 roku na konferencji naukowej „Technologiczna utopia czy dyktat Technopolu – rozważania o technologicznych zależnościach ponowoczesnych społeczeństw” zorganizowanej przez Wydział Nauk Społecznych Uniwersytetu Śląskiego w Katowicach. Przedmiotem referatu był dyskurs komentarzy internetowych dotyczących bionicznych protez dla psów. Internetowe zbiórki pieniędzy i aukcje charytatywne na rzecz finansowania operacji zwierząt były ściśle strzeżone przez administratorów, a wypowiedzi podające w wątpliwość zasadność podejmowanych działań niezwłocznie usuwano, nie podejmując nawet prób dyskusowania z adwersarzami.

⁴⁸ Jeżeli użytkownik Facebooka był gościem wirtualnego wydarzenia lub członkiem grupy, lecz został zablokowany przez administratora (lub usunął się sam), jego komentarze nie zostają automatycznie skasowane, a nazwa jego profilu wyświetlana jest w innym kolorze.

Podsumowanie

Internet stał się tkanką naszego życia – konstatuje Manuel Castells⁴⁹. Tym samym badania społeczne z konieczności zainteresowały się cyberprzeustrzeżeniem, która niesie z sobą ogromny potencjał jako pole i narzędzie badawcze. Internetowe komentarze jako wypowiedzi powstające w sytuacjach naturalnych stanowią nietuzinkowy, ale też niełatwy materiał do analizy. Szczególnie ważny jest tutaj dobór próby – czas i miejsce oraz okres gromadzenia danych. W odpowiedzi na informacyjną nadpodaż współczesnej kultury powstały mechanizmy analizy *big data*, jednak nie są one właściwe mniejszym zbiorom danym, z którymi przychodzi się mierzyć netnografom czy badaczom dyskursów. Badacz, który decyduje się na analizę komentarzy, musi zwrócić szczególną uwagę na fakt, że pracuje na danych zmieniających się w czasie rzeczywistym – danych stających się.

Nie bez znaczenia jest tu również kwestia społeczno-demograficznej charakterystyki osób wypowiadających się w komentarzach. Ich wieku, płci czy wykształcenia nie sposób określić – co więcej, współcześnie przeciętny badacz nie może nawet ocenić, czy materiał badawczy nie został stworzony przez zaprogramowaną do komentowania maszynę. Warto mieć te kwestie na uwadze przy doborze materiału do analizy: dla jednych problemów badawczych będą one znamienne, dla innych zaś wtórne.

Analiza komentarzy może obejmować dane wywoływane zarówno przez badacza, jak i przez osoby trzecie. Ingerencja badającego w materiał może rodzić wątpliwości co do naturalności wypowiedzi przez niego sprowokowanych, warto zatem rozważyć odłączenie tak uzyskanych danych od tych, które się jedynie zastało, a wyniki ich odrębnych omówień zestawić ze sobą.

Główny problem w analizie internetowych komentarzy stanowi kwestia możliwie nieskończonych przemian materiału badawczego pod względem zarówno ilościowym, jak i jakościowym. Zaprzeszanie pobierania danych konieczne jest w celu podjęcia dalszego etapu badań, lecz badacz musi mieć świadomość, że zbiór, na którym pracuje, został przez niego sztucznie zamknięty, a materiał zmienia się nieustannie. Sam przyrost danych w badaniach mediów stanowi *nihil novi*, czego przykładem jest choćby regularne wydawanie drukowanej prasy (a tej nie bada się dopiero po zamknięciu danego tytułu). W badaniach Internetu problem ten okazuje się jeszcze szerszy, czego szczególnym przykładem są komentarze na stronach www. Zagadnienie edycji oraz możliwości oddolnego i odgórnego utracenia da-

⁴⁹ M. Castells, *Galaktyka Internetu. Refleksje nad Internetem, biznesem i społeczeństwem*, tłum. T. Hornowski, Warszawa 2003, s. 11.

nych jest nowym wyzwaniem dla badaczy mediów. Zapisywanie zrzutów ekranu pozwala zaobserwować zmiany w ilości i jakości komentarzy, niemniej wykorzystywanie materiału, który został zmieniony lub usunięty, budzi etyczne kontrowersje. Ilościowa analiza fluktuacji danych wydaje się aksjologicznie neutralna, gdyż podnosi wiarygodność badania. Wymaga ona jednak wzmożonej pracy nad zebraniem materiałem.

THE PROBLEM OF DATA FLUCTUATION IN RESEARCH CARRIED OUT IN CYBERSPACE – THE CASUS OF ONLINE COMMENTS

ABSTRACT

The article deals with the issue of Internet research methodology in the culture of distraction and informational oversupply. Contemporary internet sociology is facing not only the uncontrolled data growth, but also their variability, which takes place in real time. The work discusses the phenomenon of the analysis of online comments, where the analysis of data fluctuation gives information on the mechanisms of communication in cyberspace. At the same time, the ethical dilemmas of using this research material have been raised.

KEYWORDS

big data, netnography, variable data, becoming data, internet research

BIBLIOGRAFIA

1. Babbie E., *Badania społeczne w praktyce*, tłum. W. Betkiewicz et al., Warszawa 2006.
2. Cartwright D. P., *Zastosowania analizy treści*, [w:] *Metody badań socjologicznych*, oprac. S. Nowak, Warszawa 1965.
3. Castells M., *Galaktyka Internetu. Refleksje nad Internetem, biznesem i społeczeństwem*, tłum. T. Hornowski, Warszawa 2003.
4. Cyrek B., *Media społecznościowe – nowa przestrzeń nauki*, „Kognitywistyka i Media w Edukacji” 2016, nr 2.
5. *Debating Deliberative Democracy*, eds. J. Fishkin, P. Laslett, Malden 2006.
6. Garncarek P., *Teksty kultury w nauczaniu języka polskiego jako obcego*, „Acta Universitatis Lodziensis: Kształcenie Polonistyczne Cudzoziemców” 2010, nr 17.
7. Hammersley M., Atkinson P., *Metody badań terenowych*, tłum. S. Dymczyk, Poznań 2000.
8. Jemielniak D., *Netnografia, czyli etnografia wirtualna – nowa forma badań etnograficznych*, „Prakseologia” 2013, nr 154.
9. Jones S. R. G., *Was There a Hawthorne Effect?*, [online] <https://www.economics.mcmaster.ca/research/publications/working-papers/1985-1999/91-01.pdf> [dostęp: 7.03.2018].
10. Jurek K., *Badania społeczne w internecie. Wirtualna etnografia w teorii i praktyce*, „Nauka i Szkolnictwo Wyższe” 2013, nr 1 (41).

11. *Kodeks etyki socjologa*, [online] <http://pts.org.pl/wp-content/uploads/2016/04/kodeks.pdf> [dostęp: 9.03.2018].
12. Krajewski M., *Kultura dystrakcji – deficyty uwagi i strategie jej kumulacji*, [online] http://krajewskimarek.blox.pl/resource/Ekonomia_uwagi.doc [dostęp: 3.01.2017].
13. Lamme V. A., *Why Visual Attention and Awareness Are Different*, "Trends in Cognitive Sciences" 2003, No. 7 (1).
14. Lau H. C., Rosenthal D., *Empirical Support for Higher-order Theories of Conscious Awareness*, "Trends in Cognitive Sciences" 2011, No. 15 (8).
15. Lisowska-Magdziarz M., *Analiza tekstu w dyskursie medialnym: Przewodnik dla studentów*, Kraków 2007.
16. Lisowska-Magdziarz M., *Analiza zawartości mediów. Przewodnik dla studentów*, Kraków 2004.
17. Mach-Król M., *Analiza i strategia big data w organizacjach*, "Studies & Proceedings of Polish Association for Knowledge Management" 2015, Vol. 74.
18. McCombs M. E., Shaw D. L., Weaver D. H., *New Directions in Agenda-setting Theory and Research*, "Mass Communication and Society" 2014, Vol. 17, No. 6.
19. Mitchell W. J. T., *Pokazując widzenie: krytyka kultury wizualnej*, „Artium Quaestiones” 2006, nr XVII.
20. Moolenaar N. M., *A Social Network Perspective on Teacher Collaboration in Schools: Theory, Methodology and Applications*, "American Journal of Education" 2012, No. 1.
21. Prensky M., *Digital Natives, Digital Immigrants*, [online] <https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf> [dostęp: 1.12.2017].
22. Rygielska M., *O „tekście kultury”*, „Zeszyty Etnologii Wrocławskiej” 2015, nr 1.
23. Simon H., *Designing Organizations for an Information-Rich World*, [in:] *Computers, Communications, and the Public Interest*, ed. M. Greenberger, Baltimore 1971.
24. Szczepaniak K., *Zastosowanie analizy treści w badaniach artykułów prasowych – refleksje metodologiczne*, „Acta Universitas Lodziensis: Folia Sociologica” 2012, t. 42.
25. Szlendak T., *Kultura nadmiaru w czasach niedomiaru*, „Kultura Współczesna” 2013, nr 1.
26. Szpunar M., *Badania Internetu vs. badania w Internecie, czyli jak badać nowe medium – podstawowe problemy metodologiczne*, „Studia Medioznawcze” 2007, nr 2 (29).
27. Szpunar M., *Imperializm kulturowy Internetu*, Kraków 2017.
28. Szpunar M., *Internet w procesie gromadzenia danych o charakterze wtórnym*, „Studia Medioznawcze” 2009, nr 2 (37).
29. Szpunar M., *Nauka przeciw fałszerstwom?*, [w:] *Człowiek – media – edukacja*, red. J. Morbitzer, E. Musiał, Kraków 2013.
30. Szpunar M., *Nowe-stare medium. Internet między tworzeniem nowych modeli komunikacyjnych a reprodukowaniem schematów komunikowania masowego*, Warszawa 2012.
31. Szpunar M., *Wrażliwość (nie tylko) artystyczna*, „Kultura i Społeczeństwo” 2017, nr 1.
32. Tabakow M., Korczak J., Franczyk B., *Big Data – definicje, wyzwania i technologie informatyczne*, „Informatyka ekonomiczna” 2014, nr 1 (31).
33. Wasserman S., Faust K., *Social Network Analysis: Methods and Applications*, Cambridge 2009.
34. Welsch W., *Estetyka i anestetyka*, tłum. M. Łukasiewicz, [w:] *Postmodernizm. Antologia przekładów*, red. R. Nycz, Kraków 1998.

35. Wierzchoń M., Gruszka A., *Czy uwaga ekstensywna jest świadoma? O relacjach uwagi i świadomości w kontekście teorii stanów uwagi*, „Studia Psychologiczne” 2011, t. 49, z. 3.
36. Wójcik M., *Big data w zarządzaniu informacją – przegląd wybranych zagadnień*, [w:] *Inspiracje i innowacje: zarządzanie informacją w perspektywie bibliologii i informatologii*, red. S. Cisek, Kraków 2016.

