

PRZEGLĄD  

---

SOCJOLOGICZNY

tom LXI/1  

---

2012



---

ŁÓDZKIE TOWARZYSTWO NAUKOWE

---

PRZEGLĄD  

---

SOCJOLOGICZNY

„Pragniemy (...) zmanifestować, że pismo nie ma wyrażać poglądów jakiegoś «kierunku» czy «szkoły», lecz reprezentować polski ruch socjologiczny w ogóle”.

*(Florian Znaniecki w liście do Ludwika Krzywickiego, w związku z przygotowaniem I tomu „Przeglądu Socjologicznego”, październik, 1929 r.)*

PRZEGLĄD  

---

SOCJOLOGICZNY

tom LXI/1  

---

2012



---

ŁÓDZKIE TOWARZYSTWO NAUKOWE

---

REDAKCJA NACZELNA WYDAWNICTW  
ŁÓDZKIEGO TOWARZYSTWA NAUKOWEGO  
**A. Sławomir Gala, Edward Karasiński,**  
**W. Małgorzata Krajewska** (redaktor naczelny), **Jan Szymczak**  
90-505 Łódź, ul. M. Skłodowskiej-Curie 11  
tel. (42) 665-54-59; fax: (42) 665-54-64  
Sprzedaż wydawnictw: (42) 66-55-448, <http://sklep.ltn.lodz.pl>  
<http://www.ltn.lodz.pl> e-mail: [biuro@ltn.lodz.pl](mailto:biuro@ltn.lodz.pl)

RADA REDAKCYJNA

**Krzysztof Gorlach, Zygmunt Gostkowski, Władysław Markiewicz, Harri Melin,**  
**Fritz Schütze, Kazimierz M. Słomczyński, Antoni Sulek, Piotr Sztompka,**  
**Włodzimierz Wesolowski, Włodzimierz Winclawski, Marek Ziółkowski**

KOMITET REDAKCYJNY

**Zbigniew Bokszański, Marek Czyżewski, Kaja Kaźmierska,**  
**Jolanta Kulpińska** – redaktor naczelny,  
**Krystyna Lutyńska, Paweł Starosta,**  
**Wielisława Warzywoda-Kruszyńska**

REDAKTOR TOMU: **Anna Kubiak**  
SEKRETARZ REDAKCJI: **Iwona Kociemska**  
REDAKTOR JĘZYKOWY: **Joanna Balcerak**

RECENZENCI ZEWNĘTRZNI

**Krzysztof Czekaj, Krystyna Janicka, Katarzyna Kaniowska, Andrzej Piotrowski**  
**Andrzej Sadowski, Lynda Walters, Włodzimierz Winclawski**

Wydano z pomocą finansową Ministerstwa Nauki i Szkolnictwa Wyższego,  
Dziekana Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego

Copyright by Łódzkie Towarzystwo Naukowe – Łódź 2012  
ISSN 0033-2356

Wydanie I, wersja drukowana pierwotna

PROJEKT OKŁADKI - Hanna Stańska  
OPRACOWANIE KOMPUTEROWE: „PERFECT”, tel. (42) 215-83-46  
DRUK: „GRAFIX”, 90-644 Łódź, ul. Żeligowskiego 46, tel./fax: (42) 651-96-35  
[www.grafix-poligrafia.pl](http://www.grafix-poligrafia.pl) e-mail: [grafix@grafix-poligrafia.pl](mailto:grafix@grafix-poligrafia.pl)

NAKLAD: 250 egz.

## SPIS TREŚCI

|  |     |
|--|-----|
| Od redakcji .....  | 7   |
| <b>ARTYKUŁY</b>  |     |
| Anna Kubiak, Aneta Krzewińska – Sondaż deliberatywny – inwentarz problemów ....  | 9   |
| Franciszek Sztabiński, Teresa Żmijewska - Jędrzejczyk – <i>Mixed Mode Survey</i><br><i>Design: problem efektu techniki</i> .....                     | 3   |
| Rafał Boguszewski, Natalia Hipisz – Od kwestionariusza papierowego<br>do wspomaganego komputerowo. Analiza porównawcza technik PAPI i CAPI .....     | 65  |
| Katarzyna Grzeszkiewicz-Radulska – Jak w Polsce rozwija się zjawisko<br>nieodpowiedzi respondentów? Analiza na przykładzie badań CBOS 1993–2011..... | 83  |
| Natalia Hipisz – Standaryzowane techniki badawcze jako szczególny sposób<br>komunikowania się .....  | 113 |
| Wojciech Jabłoński – Sytuacje trudne w wywiadzie telefonicznym .....   | 133 |
| Sylwia Męćfał – Problemy badań terenowych – wybrane kwestie metodologiczne,<br>praktyczne oraz etyczne przy badaniu zjawisk „trudnych”.....          | 155 |
| Rafał Maciąg – Społeczna konstrukcja dokumentów i danych urzędowych.<br>Dane medyczne .....  | 179 |
| Jolanta Lisiek-Michalska – Ocena metodologicznej i praktycznej jakości fokusa<br>– propozycja procedury .....  | 201 |
| <b>RECENZJA</b>  |     |
| Klaus Krippendorff, Content Analysis. An Introduction to Its Methodology<br>Thousand Oaks, CA: Sage Publications, 2004 (Sylwia Męćfał) .....         | 229 |

## LIST OF CONTENTS

|  |    |
|--|----|
| <b>ARTICLES</b>  |    |
| Anna Kubiak, Aneta Krzewińska – Deliberative polling® – a list of problems .....   | 30 |
| Franciszek Sztabiński, Teresa Żmijewska - Jędrzejczyk – <i>Mixed Mode Survey</i><br><i>Design: Mode Effect Problem</i> .....           | 63 |
| Rafał Boguszewski, Natalia Hipisz – From the paperquestionnaire<br>to the computer-assisted one. The comparison of PAPI and CAPI ..... | 82 |

---

|   |     |
|---|-----|
| Katarzyna Grzeszkiewicz-Radulska – How is the nonresponse phenomenon developing in Poland? A trend analysis of nonresponse in Poland in 1993–2011 based on CBOS research .....                        | 111 |
| Natalia Hipisz – Standardized research techniques as a special type of communication – the comparison of the personal standardized interview, the telephone interview and the auditorium survey ..... | 131 |
| Wojciech Jabłoński – Difficult situations in telephone interviews .....   | 153 |
| Sylwia Męćfał – Problems in field work/field studies – chosen methodological, practical and ethical issues in exploring “difficult” phenomena .....   | 178 |
| Rafał Maciąg – The social construction of official documents and data. Medical data.....  | 199 |
| Jolanta Lisiek-Michalska – Methodological and practical evaluation of fokus group interview (FGI) – a proposal of evaluative procedure .....  | 227 |

## OD REDAKCJI

Niniejszy tom „Przeglądu Socjologicznego” poświęcony jest problematyce warsztatu badawczego. Zawiera teksty dotyczące stosowania różnorodnych technik badawczych, ich ograniczeń i modyfikacji. Autorzy – przedstawiciele różnych ośrodków akademickich i firm badawczych – rozważają problemy związane z różnorodnymi zastosowaniami badań sondażowych:

- z wykorzystaniem i ograniczeniami sondażu deliberatywnego – stosowaniem procedury *Mixed Mode*, a więc łączenia w badaniu zarówno różnych technik wywiadu, jak i ankiety, oraz zjawiska niedostępności respondentów;
- specyfiki komunikacyjnej w wywiadach osobistych, telefonicznych i wspomaganych komputerowo.

Prezentujemy również analizy dotyczące wykorzystywania danych urzędowych i badań fokusowych, a także rozważania dotyczące zjawisk „trudnych” w badaniach terenowych.



ANNA KUBIAK\*  
ANETA KRZEWIŃSKA\*\*  
Uniwersytet Łódzki

## SONDAŻ DELIBERATYWNY® – INWENTARZ PROBLEMÓW

### Streszczenie

W artykule prezentujemy sondaż deliberatywny® – metodę badania opinii, stworzoną i opatentowaną przez Jamesa S. Fishkina. W pierwszej części opisujemy poszczególne etapy prowadzenia tego typu badań oraz prezentujemy sondaż, zrealizowany w Poznaniu w 2009 r. (jedyne przeprowadzone w Polsce), który dotyczył poglądów mieszkańców miasta na temat zarządzania stadionem po Mistrzostwach Europy w Piłce Nożnej EURO 2012. W drugiej części artykułu przedstawiamy wielowątkową metodologiczną analizę tej metody badawczej, wskazując na problemy związane z realizacją takiego badania, a także podkreślając zalety tego przedsięwzięcia, do których zaliczamy przede wszystkim aktywizowanie obywateli i włączanie ich w proces podejmowania decyzji.

**Słowa kluczowe:** sondaż deliberatywny®, demokracja deliberatywna, deliberacja

### WSTĘP

Sondaż deliberatywny®<sup>1</sup> jako metoda i procedura badawcza został opracowany, przedstawiony i opatentowany przez Jamesa S. Fishkina<sup>2</sup>. Jest to wynik jego przemyśleń, zawartych w książce *Democracy and Deliberation: New Di-*

\*anna.l.kubiak@neostrada.pl \*\*anetak@uni.lodz.pl

<sup>1</sup> Sondaż deliberatywny® jest chronioną patentem autorską metodą badawczą profesora Jamesa S. Fishkina i jego nazwa powinna posiadać odpowiednie oznaczenie „®”. Badania z wykorzystaniem tej metody można przeprowadzić tylko za zgodą jej twórcy, a w trakcie realizacji powinien być obecny J.S. Fishkin i jego zespół badawczy.

<sup>2</sup> James S. Fishkin jest profesorem w Departamencie Komunikacji Uniwersytetu Stanforda oraz dyrektorem Centrum Demokracji Deliberatywnej.

wyborów w sferze polityki, ludzie często stosują procedury znane im ze sfery konsumpcji. „Sposób, w jaki kształtowani są przez współczesne społeczeństwo jego obywatele, podporządkowany jest przede wszystkim roli konsumenta, jaką mają do odegrania” [Bauman 2000: 95].

Próba przeciwdziałania takiemu „konsumenckiemu” podejściu do spraw istotnych społecznie została podjęta przez przedstawicieli demokracji deliberatywnej<sup>5</sup>, przy czym nie chodzi im o prosty powrót do demokracji bezpośredniej lub uczestniczącej, ale raczej o znalezienie środków czy procedur, wzmacniających poziom partycypacji i kształtujących kompetencje obywatelskie. Demokracja deliberatywna to „[...] każda z koncepcji, w których zakłada się, że namysł wolnych i równych obywateli w sprawach publicznych stanowi rdzeń prawomocnego podejmowania decyzji politycznych oraz samorządności” [Bohman 1998: 401]. Prawomocność polityczna jest więc związana nie tylko z wrzuceniem kartki do urny wyborczej i z zastosowaniem czystej zasady większości, lecz także z podawaniem argumentów, racji i wyjaśnień dla czynionych rozstrzygnięć [Held 2010: 307]. Zdaniem teoretyków demokracji deliberatywnej, właśnie dzięki namysłowi (*deliberation*) można przewyciężyć ograniczenia charakterystyczne dla poglądów indywidualnych, prywatnych i podnieść jakość procesu publicznego podejmowania decyzji. Taki publiczny namysł charakteryzuje się trzema ważnymi cechami:

1) może przebudowywać świadomość jednostek i ułatwiać im rozumienie złożonych problemów świata publicznego;

2) może ujawniać „ideologiczny aspekt” pewnych preferencji, a więc ich jednostronność, wrywkowość, partykularyzm; deliberacja – odsłaniając podłoże tych partykularnych stanowisk czy preferencji, a także ich ideologiczne zniekształcenia – pozwala na dynamiczne kształtowanie opinii publicznej;

3) może zastąpić „język interesu językiem rozumu” [Elster 1989: 111]. Deliberacja nie zatrzymuje się tylko na upowszechnianiu informacji i wymianie poglądów, umożliwia ich analizę oraz weryfikację wysuwanych argumentów, a tym samym udoskonala sądy wypracowywane zbiorowo. J. Fishkin uważa, że

---

<sup>5</sup> Chociaż pojęcia *demokracja deliberatywna* używają reprezentanci różnych stanowisk, to jednak zgodnie wymienia się Jürgena Habermasa i Johna Rawlsa wśród najbardziej znanych teoretyków tego podejścia. Poglębiającą analizę porównawczą zapatrywać na demokrację deliberatywną tych dwóch filozofów znajdziemy w artykule [Bukusiński 2002: 347–363] – np. obaj filozofowie tworzą konstruktywistyczne teorie dotyczące społeczno-politycznej rzeczywistości; przypisują duże znaczenie rozumowi deliberacyjnemu w życiu politycznym i publicznym; podkreślają rolę dyskusji, dialogu, komunikacji i dyskursu itd.

*rections for Democratic Reform* (1991), dotyczących współczesnej demokracji i jej deficytów, a także próba znalezienia praktycznych rozwiązań i procedur badawczych, które mogłyby diagnozować lub ukazywać słabości współczesnej demokracji. Zdaniem J. Fishkina, współczesna demokracja narzuca wybór pomiędzy „[...] równością polityczną i stosunkowo niską kompetencją mas a nierównością polityczną i stosunkowo wysoką kompetencją elit” [Fishkin 1991: 1]. Stąd debata polityczna we współczesnych krajach demokratycznych jest na ogół powierzchowna, ujawnia brak zaangażowania i wysoki poziom wyobcowania wyborców. Badani, odpowiadając na pytania w sondażach – czy to dotyczących spraw bieżących, czy to prezentujących ich poglądy i preferencje polityczne – często posługują się tzw. racjonalną ignorancją<sup>3</sup>. Oznacza to, że nie wkładają wysiłku w poszukiwanie i przetwarzanie informacji np. na temat kandydatów startujących w wyborach, istotnych problemów ekonomicznych kraju itp., gdyż jest to zbyt długotrwały i pracochłonny proces. Zamiast tego wykorzystują informacje najłatwiej dostępne (np. te, które często są prezentowane przez media) do formułowania odpowiedzi na pytania z kwestionariuszy i podejmowania decyzji przy urnach wyborczych. Można stwierdzić, że racjonalne wykorzystywanie własnej ignorancji zmniejsza wpływ obywateli na zjawiska polityczne w ich kraju (być może bez istnienia tego mechanizmu nie byłibyśmy w stanie podjąć żadnej decyzji w złożonych sprawach publicznych). Także elity oddalają się od problemów prawdziwej, autentycznej debaty politycznej, w ich działaniach coraz częściej dominują chwytły marketingowe, a najważniejszymi instrumentami uprawiania polityki stają się sondaże opinii publicznej oraz badania fokusowe, wypierające pogłębione analizy postaw różnych grup wyborców<sup>4</sup>. Nad debatą polityczną przeważa medializacja dyskursu, utrudniająca rzeczywiste rozumienie wzajemnych interesów i preferencji wyborców.

Warto zauważyć, że we współczesnej kulturze dominuje indywidualizm decyzji „konsumenckich”, który jest sprzeczny z myśleniem „wspólnotowym” – niezbędnym w polityce rozumianej jako działanie dla dobra wspólnego. „Rozwiązywanie problemów w myśl zasad procedury demokratycznej zdaje się więc tracić na znaczeniu na rzecz rozwiązań prowadzonych zgodnie z trybem korporacyjnym, wolny rynek zaś, wraz z silnie wzorcotwórczą rolą reklamy, przeobraża obywatela w konsumenta” [Wnuk-Lipiński 2004: 105]. Rola konsumenta zaczyna dominować nad rolą obywatela w takim sensie, że dokonując

<sup>3</sup> Termin „racjonalna ignorancja” wprowadził [Downs 1965].

<sup>4</sup> O problemie praktycznego wykorzystywania wyników sondaży opinii publicznej pisał m.in. [Sułek 2001: 213–235].

w toku publicznego namysłu „[...] jego uczestnicy muszą być gotowi rozważyć istotną jakość wysuwanych tez” [Fishkin 1991: 37].

W teoriach demokracji deliberatywnej preferencje obywateli nie są traktowane jako z góry dane elementy życia politycznego, ale zakłada się, że obywatelom zostaną dostarczone narzędzia badania opinii na temat wspólnych problemów. „Chodzi tu o zaszczepienie w życiu politycznym procesu namysłu” [Held 2010: 309], umożliwiającego rozważanie poglądów zgodnie z zasadami bezstronności oraz otwartości na różne punkty widzenia i swobodnej ich prezentacji. Obywatele, przedstawiając własne racje, są jednocześnie zobowiązani do akceptacji opinii ich oponentów, bowiem dyskusje są regulowane „[...] ściśle określonymi wymogami normatywnymi”<sup>6</sup> [Wesołowska 2010: 13]. Wszystkie istotne kwestie omawiane są na forum publicznym, nie istnieją informacje tajne, poufne, dostępne tylko wybranej grupie dyskutujących. Takie deliberatywne dyskusje są najlepszą formą podejmowania decyzji w sytuacji, gdy mamy do czynienia z ograniczonymi zasobami i z góry wiadomo, że nie ma możliwości zaspokojenia potrzeb każdej ze stron konfliktu. Elżbieta Wesołowska uważa, iż dyskusje oparte na zasadach deliberacyjnych są do tego stopnia uniwersalnymi sposobami wypracowywania decyzji, że mogą być wykorzystywane „[...] w razie kontrowersji natury moralnej, a nawet w przypadkach długotrwałych, zaognionych i nierozwiązywalnych dotychczasowymi środkami konfliktów” [Wesołowska 2010: 28].

Aby możliwe było prowadzenie dyskusji w „deliberatywnym duchu”, należy spełnić kilka warunków: uczestnicy winni dysponować elementarnymi kompetencjami wynikającymi z zapoznania się z informacjami dotyczącymi danego zagadnienia, muszą sformułować własne stanowiska na podstawie analizy danego problemu i znaleźć argumenty wspierające je, wreszcie potrzebna jest otwarta, obywatelska debata, w której różne propozycje mają szansę w równym stopniu zostać zaprezentowane [Juchacz 2006: 17]. Dyskusja powinna być wolna od jakichkolwiek przymusów, musi angażować wszystkich potencjalnie związanych z danym problemem uczestników, którzy będą dyskutować, nie zapominając także o tych, których kwestie te dotyczą tylko pośrednio i którzy nie są obecni podczas

---

<sup>6</sup> Zasady prowadzenia dyskusji deliberacyjnych są tworzone przed rozpoczęciem dyskusji i przedstawiane ich uczestnikom na początku spotkania. Katalog tych zasad bywa różny, ale do najważniejszych należy zaliczyć: odnoszenie się do siebie ze wzajemnym szacunkiem; traktowanie innych osób w grupie jak równoprawnych partnerów; podkreślanie, iż każda opinia jest jednakowo ważna, a każdy głos ma zostać wysłuchany; szczerym i subiektywnym opiniom powinny towarzyszyć odpowiednie uzasadnienia; pojedyncze sądy należy wyrażać w sposób zrozumiały dla pozostałych uczestników dyskusji; należy słuchać innych, nie przerywać im wypowiedzi itp. [Reykowski 2007: 216].

debaty. „W trakcie takiej dyskusji może następować przekształcenie preferencji uczestników, zmiany w sposobie uzasadniania tych preferencji lub ich hierarchicznej organizacji, a także poznanie i zrozumienie stanowisk odmiennych niż własne. Ostatecznym efektem tych modyfikacji może być wypracowanie stanowiska grupowego możliwego do zaakceptowania przez wszystkich (lub przynajmniej zdecydowaną większość) zainteresowanych” [Wesołowska 2010: 31]. Czasami nie można sformułować wniosków, akceptowanych przez wszystkich uczestników lub chociażby ich większość, wtedy równie cenne jest określenie kwestii, w których udało się osiągnąć porozumienie oraz wskazanie ewentualnych „obszarów niezgody”, dla których trudno było wypracować kompromis.

Wartością dodaną deliberacji jest swoiste „uczenie się obywatelskości” [Wesołowska 2010: 31], związane z nabywaniem kompetencji pomagających jednostce funkcjonować w systemie demokratycznym, zarówno na poziomie lokalnym (np. własnej wspólnoty lokalnej), jak i globalnym (deliberacje obywateli Unii Europejskiej). Nabyte umiejętności mogą być przez deliberujących przenoszone na inne sfery ich życia. Uczestnictwo w takiej dyskusji powinno „[...] stymulować rozwój poznawczy, zwiększać kompetencje komunikacyjne, poprawiać zdolności kooperacyjnego rozwiązywania problemów i uczyć myślenia w kategoriach wspólnego dobra i sprawiedliwości społecznej” [Wesołowska 2010: 31]. E. Wesołowska pisze, że osoba, której uda się dojść do porozumienia z innymi uczestnikami dyskusji odnosi osobisty sukces, jest przekonana o własnej skuteczności i sile prezentowanych przez siebie argumentów.

Na szczególną uwagę zasługuje także fakt, iż wyniki deliberacji mogą zostać wykorzystane na kilka sposobów. Wnioski z dyskusji odgrywają niekiedy rolę edukacyjną, kiedy decydenci nie są informowani ani o tym, że miała miejsce deliberacja, ani o tym, jakie decyzje wypracowano podczas jej trwania. Samo deliberowanie jest wtedy korzystne tylko dla biorących w nim udział, umożliwia rozwijanie umiejętności związanych z lepszym funkcjonowaniem w systemie demokratycznym. Ponadto deliberacje mogą pełnić funkcję konsultacyjną – wnioski z dyskusji mogą, ale nie muszą być wzięte pod uwagę przez podejmującą decyzję władzę, która właśnie dzięki deliberacjom ma okazję zapoznać się z opiniami obywateli na dany temat. Deliberacja może być także „[...] rzeczywistym procesem podejmowania wiążących decyzji” [Wesołowska 2010: 30]. W takiej sytuacji decydenci realizują to, co zostało ustalone przez deliberujących obywateli. W wersji mniej kategorycznej przyjmuje się, że pomysły wypracowane przez grupę nie muszą znaleźć swojego przełożenia na działanie, ale zawsze podejmujący decyzję powinien wyjaśnić, z jakiego powodu odstąpił od wprowadzenia wypracowanych rozwiązań.

### SONDAŻ DELIBERATYWNY®

Wprowadzenie w życie – aplikacja reguł sformułowanych przez teoretyków demokracji deliberatywnej – wymaga odpowiednich metod, procedur badawczych. Jedną z propozycji jest sondaż deliberacyjny® Jamesa S. Fishkina, który po raz pierwszy został zorganizowany w roku 1994. Do tej pory przeprowadzono ponad 20 takich sondaży dotyczących m.in.: przestępczości (Wielka Brytania 1994), utrzymania monarchii (Australia 1996), wprowadzenia euro (Dania 2000), sytuacji Aborygenów (Australia 2001), sytuacji Romów (Węgry 2001, Bułgaria 2002), systemu opieki zdrowotnej (Włochy 2006), przyszłości Europy (Unia Europejska 2007), bezrobocia (Węgry 2008), lokalnego budżetu (Chiny 2008), zarządzania miejskim stadionem sportowym (Polska 2010). Prowadzenie takiego sondażu zakłada spełnienie wielu warunków i wykonanie różnorodnych przedsięwzięć. Oto najważniejsze z nich:

- Działania organizacyjne – skompletowanie zespołu badawczego (merytorycznego, zajmującego się także niezmiernie istotnymi sprawami organizacyjnymi); pozyskanie sponsorów i współpracujących mediów; znalezienie specjalistów w dziedzinie stanowiącej przedmiot deliberacji; wybór firmy przeprowadzającej badania sondażowe; wybór miejsca przystosowanego do prowadzenia deliberacji.

- Wylosowanie reprezentatywnej próby (mieszkańców jakiegoś państwa, miasta, gminy, w zależności od tego, czego dotyczy badany problem, co jest przedmiotem deliberacji) i zrealizowanie badania opinii według wcześniej przygotowanego kwestionariusza. Sondaże można przeprowadzić na różne sposoby (np. wywiady *face-to-face* lub wywiady telefoniczne). Niekiedy, sondaż przeprowadzony z obywatelami nie poinformowanymi określa się nazwą „sondażu zerowego” [Bukowski 2011: 135].

- Wylosowanie podpróby spośród uczestników „sondażu zerowego”, zaproszenie ich do uczestnictwa w debacie oraz zabieganie o to, aby wszystkie osoby wylosowane do podpróby wzięły udział w dalszych fazach badania.

- Przygotowanie i dostarczenie wszystkim członkom wylosowanej podpróby tzw. zbalansowanych materiałów informacyjnych na temat przedmiotu deliberacji, w których należy przedstawić w jak najbardziej wyczerpujący, bezstronny, a jednocześnie przystępny sposób różne punkty widzenia związane z daną sprawą. Za przygotowanie materiałów odpowiedzialni są niezależni eksperci.

- Przygotowanie i przeszkolenie facylitatorów, a więc osób, które będą czuwały nad dyskusją prowadzoną w małych grupach. Grupy te tworzy się poprzez losowy podział członków podpróby. Facylitatorzy mają być osobami sprawnymi komunikacyjnie, werbalnie, przygotowanymi do przeprowadzenia dyskusji –

aktywizowania jej uczestników, przeciwdziałania dominacji niektórych osób w trakcie debaty, trzymania się tematu debaty. Reasumując – facylitator „to ktoś w bardzo wysokim stopniu bezstronny, kto, niebędąc członkiem grupy, pracuje dla jej dobra.” [Schwarz 2002:8].

– Przeprowadzenie debaty – debata może trwać kilka godzin, cały dzień, a niekiedy nawet dwa dni. Na debatę składają się: dyskusje w małych grupach i sesje plenarne. W debatach i sesjach plenarnych uczestniczą eksperci reprezentujący zainteresowane strony oraz specjaliści z zewnątrz. W trakcie sesji każdy członek grupy może zadawać ekspertom pytania dotyczące omawianych kwestii, które zostały sformułowane w trakcie dyskusji w małych grupach. Deliberacji w charakterze obserwatorów mogą się przyglądać dziennikarze i inne zainteresowane osoby, np. reprezentujące organizacje pozarządowe. Debata, jeśli uczestnicy wyrażą na to zgodę, może być rejestrowana i oglądana przez zewnętrznych obserwatorów w innym terminie.

– Efektem takiej deliberacji jest wypracowanie stanowiska w przedmiocie debaty. Dąży się do consensusu, jeśli nie jest on możliwy wskazuje się elementy, w których debatujący byli zgodni oraz ujawnia się rozbieżne stanowiska.

– Na zakończenie przeprowadza się kolejne badanie opinii (może być to badanie w postaci wywiadów kwestionariuszowych z udziałem ankietera, bądź ankiet wypełnianych samodzielnie przez uczestników), w którym biorą udział tylko uczestnicy deliberacji odpowiadając w większości na te same pytania, które zadawano im w „sondażu zerowym”.

– Wyniki drugiego sondażu<sup>7</sup> powinny być podstawą do podjęcia decyzji w przedmiocie debaty.

– O przebiegu deliberacji i jej wynikach powinno się informować w mediach.

### **Sondaż deliberatywny® w Polsce**

W Polsce sondaż deliberatywny® zastosowano, jak do tej pory, tylko raz, chociaż mieszkańcy naszego kraju brali udział w dwóch sondażach przeprowadzonych dla obywateli Unii Europejskiej<sup>8</sup>. Polski sondaż deliberatywny® przeprowadzono w Poznaniu w listopadzie 2009 r. i dotyczył on poglądów mieszkańców na temat zarządzania stadionem miejskim. Istniejący stadion, znajdujący się przy ulicy Bułgarskiej w Poznaniu należał do miasta, był finansowany z publicznych

<sup>7</sup> Oczywiście badacze muszą zakodować wyniki, założyć bazy danych i wykonać odpowiednie obliczenia.

<sup>8</sup> Do tej pory zostały zorganizowane dwa sondaże deliberatywne® przeprowadzone z obywatelami z 27 krajów należących do Unii Europejskiej: w 2007 r. – *Europa Jutra* i w 2009 *EuroPolis*.

środków, a jego głównym użytkownikiem był klub piłkarski „Lech”. Po przebudowie stadionu na potrzeby Mistrzostw Europy w Piłce Nożnej w 2012 r., wydatki związane z utrzymaniem infrastruktury znacznie wzrosną, stąd też pojawiły się pytania, w jaki sposób go finansować, a co za tym idzie, kto powinien zarządzać obiektem i komu, przede wszystkim, miałby on służyć. Aby znaleźć odpowiedzi na te pytania zorganizowano sondaż deliberacyjny<sup>9</sup> wśród poznanian.

Część terenową badania rozpoczęto od przeprowadzenia sondażu opinii<sup>9</sup> na 880 osobowej (próbą zrealizowaną liczyła 854 mieszkańców) reprezentatywnej próbie dorosłych mieszkańców Poznania. Następnie z próby pobrano 150 osobową podpróbę (120 osób do próby zasadniczej i 30 – rezerwowej), którą zaproszono do dyskusji. Jak możemy przeczytać w dokumencie omawiającym wyniki tego badania: „Grupa uczestników deliberacji oraz grupa uczestników sondażu, którzy nie wzięli udziału w deliberacji, nie różniły się w sposób istotny statystycznie pod względem płci, poziomu wykształcenia oraz dochodów. Natomiast uczestnicy debat byli średnio nieco młodszy. [...] Dlatego też na podstawie odpowiedzi uczestników deliberacji można wnioskować o opiniach całej populacji Poznania” [Przybylska, Siu 2010: 8].

Zanim członków podpróby zaproszono na deliberację, dostarczono im materiały informacyjne<sup>10</sup>, z którymi mieli się zapoznać przed dyskusją. Broszura, którą otrzymali, liczyła 40 stron, składała się z 6 rozdziałów i była bardzo starannie zredagowana i wydana. W pierwszym z rozdziałów przedstawiono nową koncepcję stadionu, który miałby stać się obiektem wielofunkcyjnym: oprócz funkcji sportowej związanej z uprawianiem zarówno sportu wyczynowego, jak i amatorskiego, powinien pełnić również funkcje społeczne, gdyż stadion może być ośrodkiem działalności „kulturowej, społecznej, charytatywnej i rekreacyjnej” [Sondaż 2009: 5] i funkcje biznesowe, czyli powinien przynosić zysk. W rozdziale drugim zaprezentowano cztery przykłady europejskich stadionów (Stadion Miejski w Manchesterze, Stadion Amsterdam Aren A, Stadion Narodowy Francji, Stadion Olimpijski w Monachium), pisząc o genezie powstania każdego z nich, sposobach ich finansowania, ich głównej i dodatkowej działalności. W rozdziale *Sport w Poznaniu* czytelnicy mogli zapoznać się ze strukturą wydatków, które miasto ponosi na utrzymanie obiektów sportowych, z rodzajami obiektów sportowych w Poznaniu, z rozmiarami aktywności sportowej i rekreacyjnej poznanian oraz z różnorodnością imprez sportowych organizowanych w mieście.

<sup>9</sup> Badanie przeprowadziła firma Pentor metodą CATI.

<sup>10</sup> Czytelników zainteresowanych tymi zagadnieniami odsyłamy do dokumentów, które można pobrać ze strony: <http://www.ps2012.pl/index.php/Konsultacje/Sondaz-deliberatywny-w-Poznaniu>.



Kolejne rozdziały odnoszą się w bezpośredni sposób do przedmiotu deliberacji. I tak, w rozdziale czwartym zostają przybliżone: historia miejskiego stadionu, źródła finansowania jego modernizacji na EURO 2012, a także wstępne plany dotyczące późniejszego wykorzystania stadionu. Następnie czytelnik może zapoznać się z czterema alternatywnymi scenariuszami wykorzystania obiektu po zakończeniu Mistrzostw Europy w Piłce Nożnej. Każda z alternatyw – *Stadion dla Lecha, Wielkopolskie Centrum Rekreacji i Sportu, Prestiżowe centrum sportu i Bulwar – biznesowe centrum sportu i rozrywki* – została opisana za pomocą tych samych punktów, tzn.:

- wskazano instytucję zarządzającą i sposób zarządzania;
- określono stosunek kosztów do zysków;
- przedstawiono pomysł na zagospodarowanie stadionu i jego otoczenia;
- podsumowano ten opis poprzez wylistowanie plusów i minusów każdego rozwiązania.

Całość dokumentu kończy krótka notatka dotycząca metody sondażu deliberatywnego®.

Uczestnicy (w dniu deliberacji stawiło się 148 osób) przed rozpoczęciem deliberacji zostali poproszeni o wypełnienie ankiety z pytaniami, na które odpowiadali w sondażu telefonicznym, aby dokładniej śledzić zmiany ich opinii. Dyskusje w małych 15-osobowych grupach trwały od 9.00 do 18.00 i były przeplatane z dyskusjami panelowymi, w których udział brali eksperci (m.in. wiceprezes Lecha, szef Poznańskich Ośrodków Sportu i Rekreacji (POSiR), wiceprezes spółki EURO Poznań 2012, rektor Akademii Wychowania Fizycznego w Poznaniu oraz ekspert zewnętrzny z warszawskiej Akademii Wychowania Fizycznego). Całości przyglądali się dziennikarze, którzy byli tam *incognito*, zostali przedstawieni jako pomocnicy moderatorów i nie mieli prawa zadawać żadnych pytań. Po zakończeniu dyskusji uczestnicy otrzymali ankiety do samodzielnego wypełnienia na miejscu, a za udział w badaniu zapłacono im oraz dodatkowo wręczono niewielkie bonusy.

Po przeanalizowaniu wszystkich pytań, które zadano osobom biorącym udział w sondażu (trzy wersje kwestionariusza nieznacznie się od siebie różniły – pytania z trzeciego pomiaru czasami pojawiły się tylko w wywiadzie telefonicznym, lub tylko w ankiecie wypełnianej przed deliberacją) okazało się, że uległy zmianie odpowiedzi na około połowę pytań.

Najważniejsze dla władz miasta, które były współorganizatorem całego przedsięwzięcia, było poznanie opinii na temat sposobów finansowania i zarządzania stadionem. Największa zmiana nastąpiła w przypadku opinii na temat finansowania działalności stadionu – finansowanie z budżetu miasta przed deliberacją

popierało 26% uczestników, a po niej już tylko 7% badanych. O 10 punktów procentowych wzrosło poparcie dwóch różnych opinii: utrzymanie stadionu z komercyjnych usług (z 19% do 29%) oraz wariant, który zyskał największe poparcie – finansowanie przez użytkowników z niewielkim udziałem miasta (z 47% do 57%). Przed deliberacją 34% badanych przekazałoby zarządzanie stadionem instytucji typu POSiR, natomiast po deliberacji procent zwolenników tego rozwiązania spadł do 16%. Zwiększył się (z 27% do 49%) odsetek mieszkańców, którzy przekazaliby zarządzanie stadionem niezależnemu operatorowi.

W pomiarze drugim i trzecim badanym zadano również kilka pytań o wiedzę na temat samego miasta i jego obiektów sportowych. „Uśredniając stopień trafnych odpowiedzi, można powiedzieć, że po debatach wiedza uczestników konsultacji wzrosła z 68% do 75%. Największy wzrost odnotowano w przypadku odpowiedzi na pytanie o właściciela stadionu, o 14%” [Przybylska, Siu 2010: 14].

Całemu przedsięwzięciu towarzyszyła też akcja informacyjna w mediach lokalnych i krajowych. Przed sondażem mieszkańcy Poznania byli informowani o innowacyjności i przebiegu badania. Jego rezultaty zostały zaprezentowane w telewizyjnych i internetowych serwisach informacyjnych, o wynikach pisały również gazety. Prezydent Miasta – Ryszard Grobelny na łamach „Gazety Wyborczej” zadeklarował, iż wnioski z badań zostaną wzięte pod uwagę przy podejmowaniu decyzji. Zaś sami badacze i organizatorzy dość optymistycznie wypowiadali się na temat roli, jaką spełnił sondaż deliberacyjny®:

Mieszkańcy wyposażeni w wiedzę o faktach i opiniach innych niż ich własne ujrzeni przeszłość stadionu w szerszej perspektywie. Pojawiło się myślenie o stadionie w kontekście miasta i wspólnoty lokalnej. Na zagadnienie spojrzano biorąc pod uwagę wiele potrzeb, alternatywnych sposobów ich zaspokojenia, związanych z nimi plusów i minusów. W ten sposób prowadzony proces konsultacyjny prowadzi do budowania szacunku dla odmiennych racji, a także zaufania między mieszkańcami oraz między mieszkańcami i przedstawicielami władzy [Przybylska, Siu 2010: 16].

### **Sondaż deliberacyjny® a – bilans kosztów i zysków**

Warto się jednak zastanowić, czy metoda ta jest rzeczywiście remedium na bolączki demokracji przedstawicielskiej. Omawianie kwestii problematycznych związanych ze stosowaniem sondażu deliberacyjnego® zacznijmy od spraw finansowo-organizacyjnych. Należy zapłacić za możliwość przeprowadzenia takiego sondażu jego twórcy – Jamesowi S. Fishkinowi. Andrzej Bukowski opisując sondaż zorganizowany w Polsce, podaje kwotę 65 tys. dolarów opłaty

licencyjnej, od której akurat w przypadku Poznania, Fishkin odstąpił. Organizatorzy nie musieli również płacić zagranicznym ekspertom, którzy zgodzili się nie pobierać wynagrodzenia za pracę przy tym badaniu<sup>11</sup>. Z innego źródła – wywiadu z twórcą metody zamieszczonego w „Gazecie Wyborczej” – dowiadujemy się, iż należy uiścić „symboliczną opłatę” Centrum na Rzecz Deliberatywnej Demokracji. Konieczność wniesienia tej zapłaty tłumaczona jest w następujący sposób: „Pilnujemy, by eksperci reprezentowali wszystkie opcje. I by ich skład był znany na długo przed debatą – by pokazać, że żadne środowisko, żadna grupa interesów nie zostały pominięte. Za przeprowadzenie *deliberative polls*® Centrum na Rzecz Demokracji Deliberatywnej przy Uniwersytecie Stanforda pobiera symboliczną opłatę, która idzie na dalsze badania w tym zakresie. Nie chcemy, żeby gdzieś na świecie ktoś robił nieuczciwe eksperymenty korzystając z naszej nazwy” [Rozmowa 2009]<sup>12</sup>. Kwestia kosztów budzi także wątpliwości, gdyż sondaż delibeartywny® jest w naszym przekonaniu być może twórczą, ale kompilacją różnych wątków i podejść badawczych takich, jak: badania sondażowe, interwencja socjologiczna, badania fokusowe i panelowe.

Jeśli nawet uda nam się otrzymać zwolnienie z opłat za możliwość zastosowania metody, należy pamiętać, że czekają badaczy równie wysokie koszty związane z organizacją. Trzeba zgromadzić środki na przeprowadzenie reprezentatywnego badania opinii na dany temat, zapłacić niezależnym ekspertom za przygotowanie materiałów i uczestnictwo w deliberacjach, wynająć odpowiednią salę, w której zmieścimy wszystkich uczestników dyskusji, zapewnić im wyżywienie, a czasami nawet nocleg (w przypadku, gdy dyskusje w podgrupach trwają więcej niż jeden dzień, a uczestnicy deliberacji są przyjezdni). Czasami trzeba też zapłacić facylitatorom, jeśli nie zechcą pracować wolontarystycznie. I co nie jest bez znaczenia, przygotować wynagrodzenie, czy to w formie finansowej, czy to w postaci prezentów, bonusów, dla uczestników deliberacji.

Należy się zastanowić, czy gratyfikacja za uczestnictwo przewidziana dla osób badanych, nie jest swoistym „kreowaniem” zainteresowania tematem deliberacji. Odpowiednia kwota, bonusy od sponsorów, czy wycieczka do Brukseli<sup>13</sup> mogą być

<sup>11</sup> Na stronie internetowej onet.pl w notce z 24 listopada 2009 r. znajdujemy informację, że sondaż deliberatywny w Poznaniu kosztował około 120 tys. złotych.

<sup>12</sup> Informacje na temat opłaty licencyjnej są sprzeczne, gdyż 65 tys. dolarów, które jest wymienione przez Andrzeja Bukowskiego, na pewno nie jest opłatą symboliczną, a na stronie internetowej Centrum na Rzecz Demokracji Deliberatywnej przy Uniwersytecie Stanforda nie zamieszczono żadnych informacji na temat wysokości tej opłaty.

<sup>13</sup> Dwa europejskie sondaże deliberatywne® (*Europa Jutra* i *EuroPolis*) zostały zorganizowane w Brukseli i to właśnie tam zaproszono uczestników z 27 krajów, finansując ich podróż i pobyt.

silnym bodźcem do wzięcia udziału w badaniu, wypowiedziana się na dany temat, nawet, jeśli wcześniej sprawy lokalne, krajowe lub międzynarodowe zupełnie nie interesowały respondentów. Postępowanie takie może być przez niektórych traktowane jako „wątpliwe etycznie”, jak zauważa Katarzyna Grzeszkiewicz-Radulska, referując poglądy J. Goydera „[...] gratyfikacje są bardzo często formą ewidentnej manipulacji tym bardziej, że informacje uzyskane od respondenta mają znacznie większą wartość niż oferowane mu symboliczne kwoty” [Grzeszkiewicz-Radulska 2009: 108]. Z drugiej jednak strony ani badacze, ani pomysłodawcy lub organizatorzy badania nie dysponują właściwie żadnymi innymi dodatkowymi argumentami, którymi mogliby wpływać na frekwencję w tego typu przedsięwzięciach badawczych. Nie można ludzi nakłonić do udziału w deliberacjach, tak samo, jak nie jesteśmy w stanie ich przekonać do udziału w badaniach opinii, do głosowania w wyborach, do uczestnictwa w akcjach obywatelskich. Ian Shapiro uogólnia tę konstatację pisząc: „[...] wątpię, że kiedykolwiek rząd będzie w stanie faktycznie nakłonić ludzi do deliberacji. Może on przyczynić się do tego, że stanie się ona bardziej lub mniej możliwa, ale koniec końców zależy ona przede wszystkim od zaangażowania osób w niej uczestniczących. W założeniu, deliberacja wymaga dobrej woli, pełnej troski, twórczej pomysłowości i dążenia do wypracowania najlepszych rozwiązań. A tego nie można wprowadzić odgórnie” [Shapiro 2006: 66]. Czy zatem nie oczekujemy od naszych potencjalnych badanych zbyt wiele? Chcemy, aby wyrazili zgodę na udzielenie pytań w „sondażu zerowym”, a potem jeszcze zajmujemy im przynajmniej kilka godzin, które muszą spędzić wraz z innymi, nieznanymi im osobami na dyskusjach na tematy, które nie zawsze są ciekawe i mogą ich nie dotyczyć w bezpośredni sposób. Może w tej sytuacji niewielkie wynagrodzenie, jakie np. oferowano uczestnikom sondażu w Poznaniu (120 złotych rekompensaty za poświęcony czas [mm.poznan.pl 2009] i opłacenie biletu miesięcznego komunikacji miejskiej [Bukowski 2010: 142]) nie jest czynnikiem, który wpłynął na chęć udziału w tym doświadczeniu badawczym. Innym sposobem „usprawiedliwienia” płacenia za uczestnictwo jest wprowadzenie tego typu praktyk w innych badaniach społecznych, a mianowicie w zogniskowanym wywiadzie grupowym, w których zakłada się, że „[...] gratyfikacja pieniężna powinna być jedynie dodatkowym bodźcem do uczestniczenia w badaniu” [Dukaczewska-Nałęcz 1999: 154].

Na chęć wzięcia udziału w sondażu deliberatywnym<sup>®</sup>, oprócz możliwości uzyskania wynagrodzenia, ma zapewne wpływ jeszcze wiele czynników, przy czym my chcielibyśmy szczególnie zająć się jednym z nich, a mianowicie

zastanowić się nad problemem akulturacji<sup>14</sup> tej metody. Z etapem całkowitej akulturacji mamy do czynienia w sytuacji, gdy pewna procedura badacza jest „[...] społecznie uznawana za naturalną i bezpieczną [...] oraz traktowana jako powszechnie przyjęty i naukowo usankcjonowany sposób badania zjawisk społecznych” [Kubiak 2007: 63]. Czy można mówić o akulturacji sondażu deliberatywnego® traktowanego jako całościowa, zupełna procedura badawcza, czy należy zastanawiać się nad akulturacją poszczególnych jej składników? Nie dysponujemy żadnymi informacjami na temat tego, czy i co mieszkańcy Polski wiedzą na temat trzech przedsięwzięć, w których wylosowani dorośli mieszkańcy naszego kraju wzięli udział. Dysponujemy jedynie dość skąpymi doniesieniami prasowymi i internetowymi na temat poznańskiego sondażu – prezentowano wtedy nie tylko główne idee badania, ale również jego twórcę Jamesa S. Fishkina. Nie wiemy, jaka była liczba odbiorców tych komunikatów i nie wiemy czy nazwa *sondaż deliberatywny®* jest przez Polaków poprawnie kojarzona. Może należy zastanowić się, jaki jest poziom akulturacji poszczególnych części składowych sondażu deliberatywnego®? Na pewno dokonała się całkowita akulturacja badań sondażowych, możemy uznać, iż w większości mieszkańcy naszego kraju prawidłowo identyfikują role: respondenta i ankietera, wiedzą w ogólnym zarysie, na czym polega tego typu badanie i jak najczęściej prezentowane są wyniki takich badań. Gorzej przedstawia się kwestia „przygotowania” do procesu deliberowania na wskazany temat. Uczestnicy takich dyskusji powinni posiadać nie tylko odpowiednie zdolności komunikacyjne, dzięki którym będą w stanie dzielić się swoimi przemyśleniami, ale również być otwartymi na wysłuchanie, a może nawet przyjęcie argumentów innych uczestników bez przerywania, oceniania, narzucania własnej woli. Przypuśćmy, że zasady deliberacji, jeśli tylko zostaną we właściwy sposób przedstawione uczestnikom, będą przez nich przestrzegane podczas dyskusji, ale jak rozwiązać np. nierówności kompetencyjne związane z: doświadczeniami wypowiedzianego zdania w gronie nieznanymi osobami, posiadaniem wiedzy na dany temat, zainteresowaniem dyskutowanymi kwestiami? W zogniskowanych wywiadach grupowych dba się o to, aby grupa była wewnętrznie homogeniczna, dzięki temu uczestnicy np. chętniej zabierają głos i wspólnie dyskutują. W sondażach deliberatywnych® dobór do dyskutujących przy danym stoliku jest losowy, czy zatem wszyscy respondenci we własnym odczuciu są jednakowo kompetentnymi uczestnikami takiej dyskusji? Czy każdy z nich będzie równie chętnie zabierał głos w dyskusji? Czy będzie w stanie

---

<sup>14</sup> O problemie akulturacji w odniesieniu do sondażowych badań opinii pisali m.in. [Gostkowski 1966, Przybyłowska 1996].

w jasny, klarowny sposób przedstawić swoje argumenty i przekonać do nich innych? Pytania te pozostają bez odpowiedzi, ponieważ nie dysponujemy żadnymi materiałami empirycznymi, które by je rozstrzygały. „Fishkin twierdzi, że takie debaty prowadzą do otwarcia umysłów, że niektórzy wyborcy modyfikują swe stanowiska, a przede wszystkim, że nawet ludzie stosunkowo słabo wykształceni i niezamożni są w stanie pojąć skomplikowane problemy polityczne i podjąć w ich sprawie rozsądną, często kompromisową, decyzję” [Szyborski 2005]. Muszą nam wystarczyć tylko tego typu deklaracje, a twórcę tej metody wydaje się przede wszystkim interesować porównanie odpowiedzi z dwóch pomiarów opinii przeprowadzanych przed i po deliberacji, a nie sam problem dyskusji pomiędzy nimi.

Podtrzymaniu przeświadczenia (iluzji?), iż w dyskusjach uczestniczą równi sobie, bo jednakowo poinformowani obywatele, mają służyć również materiały przygotowywane i dostarczane uczestnikom przed dyskusją. Zakłada się, że są one zbalansowane, czyli przedstawiają w bezstronny i wyczerpujący sposób różne koncepcje danego zjawiska. O przygotowanie tych materiałów prosi się niezależnych ekspertów, którzy mają rozstrzygnąć, czy żadne z rozwiązań nie jest faworyzowane. Zastanawia nas to, kto dokonuje wyboru specjalistów mających przygotować takie materiały i w jaki sposób sprawdza się bezstronność ekspertów. Czy rzeczywiście istnieją materiały zbalansowane i jakimi narzędziami lub sposobami należy to „balansowanie” sprawdzać? Oczywiście jest to kolejny zbiór pytań, które pozostają bez jednoznacznej odpowiedzi, ponieważ wydaje się, iż samo czuwanie Fishkina i jego współpracowników nad doborem ekspertów do każdego sondażu nie jest żadną metodologicznie uprawomocnioną metodą postępowania.

Badani mają zapoznać się z przygotowanymi materiałami i dzięki temu dysponują podobną wiedzą, a twórca sondażu deliberatywnego zakłada, iż: po pierwsze, wylosowana podpróba rzeczywiście z uwagą przeczyta dostarczone jej informacje, będzie w stanie zrozumieć i przyswoić zamieszczone tam wiadomości na temat przedmiotu deliberacji, a ponadto będzie miała czas na zastanowienie się i rozważenie prezentowanych propozycji. Czasami badacze są w stanie to kontrolować, ponieważ zawsze mogą przed samą deliberacją przeprowadzić sondaż, w którym zadadzą szereg pytań sprawdzających wiedzę uczestników, którą mogli zdobyć po przeczytaniu dostarczonych im opracowań. Jeśli takie pytania pojawią się tylko w pomiarze przeprowadzonym po zakończeniu deliberacji, nie jesteśmy w stanie orzec, czy z wiedzą tą uczestnik przystąpił do dyskusji, czy też nabył ją w wyniku deliberacji od innych dyskutujących lub od ekspertów. Po drugie, zakładając równy stopień poinformowania badanych w danej sprawie,

przyjmuje się, że nie posiadali oni żadnej wiedzy na temat danego zjawiska przed przystąpieniem do badania i nie zainteresowali się tym problemem pod wpływem pierwszego pomiaru. Tymczasem sytuacja, kiedy uczestnicy badania opinii zostają zainspirowani do poszukiwania dodatkowych informacji przez ankieterów zadających pytania na dany temat, jest doskonale znana pod nazwą „efektu panelowego” i pojawia się dość często w badaniach panelowych<sup>15</sup> [Kosęła, Sułek 2005: 42].

Należy zastanowić się, czy rzeczywiście w przypadku sondaży deliberatywnych® osoby traktowane są bardziej podmiotowo, niż w innego rodzaju działaniach charakterystycznych dla demokracji bezpośredniej. Wszak tak samo narzucany jest temat referendum, jak temat deliberacji, czy konsultacji społecznych. Obywatele sami, oddolnie nie są w stanie zorganizować dyskusji, która byłaby wysłuchana przez decydentów, a już na pewno nie mają możliwości, aby przeprowadzić badania z wykorzystaniem omawianej tu metody badawczej. Jak pisze J. Wasilewski „Jedne kwestie są wprowadzane do deliberatywnej przestrzeni, inne są od tej przestrzeni trzymane z daleka. Na ogół nie jest to obojętne z punktu widzenia rozmaitych interesów. Jedne grupy interesów są zainteresowane w podjęciu danej kwestii, inne są zainteresowane, aby jak najdłużej zachowywała ona status *nie-deliberacji*” [Wasilewski 2006: 34]. Wybierając tematy do deliberacji możemy mieć do czynienia ze zjawiskiem sepiacji, czyli z procedurą zmierzającą do przemilczania lub aktywnego unieważniania pewnych kwestii, czy też odwrotną do niej – kontrsepiacją polegającą „[...] na uprawomocnianiu, ujawnianiu, nagłaśnianiu i czynieniu przedmiotem zainteresowania tego, co inni uznali za nieważne” [Czyżewski, Dunin, Piotrowski 2010: 19]. Możemy przypuszczać, że tematy wybierane przez osoby, które chcą sfinansować przeprowadzenie sondaży deliberatywnych® nie zawsze są najważniejszymi dla pewnej grupy społecznej uczestniczącej w badaniach, a może nawet mają odciągnąć uwagę grupy od kwestii bardziej istotnych, lecz niezbyt „wygodnych” dla decydentów. Przeglądając tematy, na które do tej pory przeprowadzono sondaże deliberatyw-

---

<sup>15</sup> Zupełnie odrębną kwestią pozostaje wykruszenie się jednostek pomiędzy pierwszym a drugim pomiarem, czyli tzw. śmiertelność grupy, kiedy wylosowani do próby respondenci nie stawiają się w dniu debaty w wyznaczonym miejscu. „Być może największą słabością jest przedmiot deliberacji. Jeżeli jest nim sprawa dotycząca wszystkich, kluczowa dla funkcjonowania całej społeczności, można spodziewać się wysokiej frekwencji. Jeśli jednak jest to kwestia drugorzędna, albo żywotnie interesująca tylko niektórych obywateli, to jaką motywację mają osoby bezpośrednio niezainteresowane, aby wziąć udział w debacie?” [Wasilewski 2006: 32]. Organizatorzy badania problemowi temu zapobiegają podobnie, jak w zogniskowanych wywiadach grupowych, zapraszając do dyskusji grupę rezerwową.

ne<sup>®</sup>, wydaje się, iż mają one raczej charakter ponadlokalny – dotyczą kwestii ważnych dla dość licznych grup. Zorganizowanie tego typu badania, np. we wspólnocie mieszkaniowej czy wśród mieszkańców jednej dzielnicy jest dość kosztowne i raczej nie wymaga reprezentowania potencjalnie zainteresowanych obywateli przez losowo dobraną grupę ich przedstawicieli. Można więc stwierdzić, iż sondaże deliberatywne<sup>®</sup> są raczej metodą do badania kwestii o szerszym, ponadlokalnym znaczeniu.

David Held formułuje tezę, iż demokracja oparta na deliberatywnych dyskusjach nie jest wcale nowatorskim modelem demokracji, ale raczej pewnym wariantem demokracji przedstawicielskiej [Held 2010: 330]. Demokracja deliberatywna w swej wzorcowej odmianie ma pozwolić na bezpośredni udział równych obywateli w podejmowaniu decyzji, a sondaż deliberatywny<sup>®</sup> oferuje uczestnictwo tylko pewnej losowo dobranej próbie. I znów w imieniu wszystkich ma się wypowiadać część, ponieważ wiele decyzji dotyczy zbyt dużych grup ludzi, aby można było ich zebrać w jednym miejscu i pozwolić wspólnie delibrować nad daną kwestią, „[...] deliberacja jako taka nie jest zajęciem dla *demos*. Nie chodzi [...] o to, że zwyczajni ludzie płci obojga są pozbawieni zdolności rozumowania, ale że 300 milionów, a nawet milion czy sto tysięcy, nie może wspólnie rozumować i snuć rozważań – jest to po prostu fizycznie niemożliwe” [Walzer 2006: 160]. A gdzie w tego typu badaniach jest miejsce dla jednostek niedostępnych, odmawiających udziału już w pierwszym pomiarze, co z osobami wykluczonymi, wyłączonymi z życia społecznego? Wszak nadal w ich imieniu wypowiadają się losowo dobrani przedstawiciele.

Przejdźmy teraz do nieco innych kwestii – James S. Fishkin jako jedną z zalet sondażu deliberatywnego<sup>®</sup> wymienia wyraźne różnice, pojawiające się w opiniach przed i po przeprowadzeniu takiego sondażu, w wywiadzie dla „Gazety Wyborczej” powiedział: „Proszę mi wierzyć, różnice w odpowiedziach przed i po są kolosalne” [Rozmowa 2009]. Może powinniśmy traktować tę deklarację jako argument na rzecz wyjątkowości i oryginalności tej metody, ale przecież w ramach klasycznego repertuaru metodologicznego, czy to w badaniach panelowych, czy w badaniach trendu, mamy różnorodne przykłady dużych zmian w opiniach, nastawieniach, zachowaniach respondentów. Ilustracją mogą być wyniki badań panelowych przedstawionych w tabeli 1.



TABELA 1. Zmiany stopnia antysemityzmu wśród tych, którzy widzieli, oraz wśród tych, którzy nie widzieli filmu *Gentlemen's Agreement* [w %]

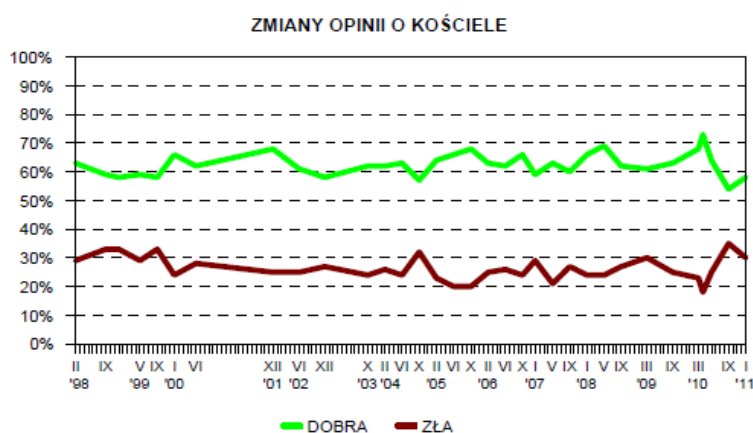
| Stopień antysemityzmu w maju następnego roku | Stopień antysemityzmu w listopadzie |             |         |             |         |             |
|--|-------------------------------------|-------------|---------|-------------|---------|-------------|
|  | wysoki                              |             | średni  |             | niski   |             |
|  | widział                             | nie widział | widział | nie widział | widział | nie widział |
| Wysoki                                       | 63                                  | 70          | 16      | 45          | 5       | 14          |
| Średni                                       | 19                                  | 15          | 46      | 20          | 9       | 16          |
| Niski  | 18                                  | 15          | 38      | 35          | 86      | 90          |
| Liczba przypadków w grupie                   | 32                                  | 132         | 26      | 76          | 57      | 173         |

Źródło: [Nowak 1965: 408].

W tabeli 1 widać np. wyraźną różnicę dotyczącą 19% osób, których stopień antysemityzmu oceniono w listopadzie jako wysoki, w maju zaś oszacowano już jako średni<sup>16</sup>.

Całkiem współczesne badania trendów pokazują, że w przypadku kwestii silnie angażujących opinię publiczną, pod wpływem nowych okoliczności, nowych informacji opinie wypowiedziane w trakcie badań sondażowych również ulegają bardzo istotnym zmianom. Ilustracją dla tej tezy mogą być wykresy 1 i 2.

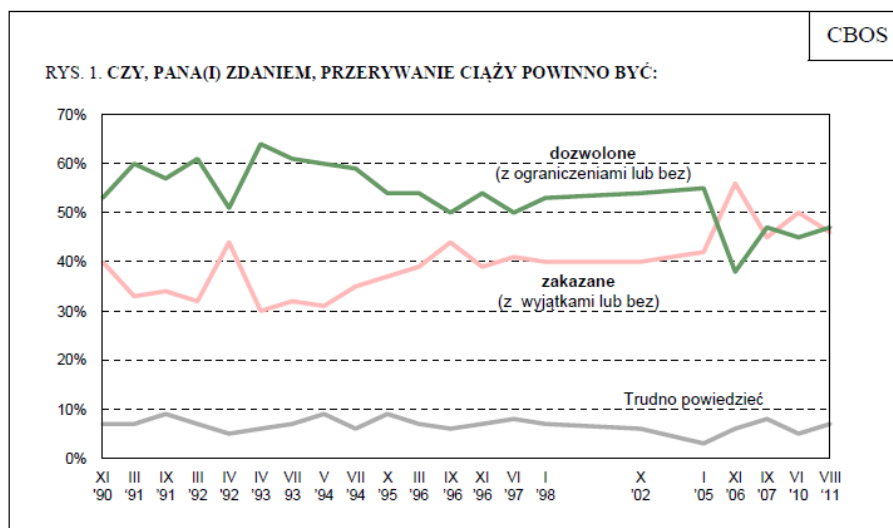
WYKRES 1. Zmiany opinii na temat Kościoła



Źródło: Komunikat CBOS, BS 8/2011. *Ocena działalności parlamentu, prezydenta i wybranych instytucji publicznych* [Feliksiak 2011: 6].

<sup>16</sup> Tabelę tę analizujemy jako tabelę panelową, w której pokazano przepływy pomiędzy poszczególnymi stopniami antysemityzmu, pomijając wpływ modyfikatora, którym w tym przypadku było obejrzenie filmu.

WYKRES 2. Zmiany opinii na temat aborcji



Źródło: Komunikat CBOS *Opinie o prawnej dopuszczalności i regulacji aborcji* [Hipsz 2011: 2].

Takie problemy, jak aborcja, stosunek do kościoła, czy inne kwestie będące przedmiotem zainteresowania opinii publicznej wywołują wyraźne zmiany w opiniach badanych biorących udział w sondażach, a różnice te czasami przekraczają nawet 20 punktów procentowych<sup>17</sup>. Taka prawidłowość nie jest charakterystyczna tylko dla polskich badań, bowiem również w badaniach prowadzonych w innej kulturze, np. w sondażu zorganizowanym w Tanzanii<sup>18</sup>, a dotyczącym profilaktyki związanej z HIV/AIDS i kontroli urodzeń, zaobserwowano bardzo wyraźne zmiany pod wpływem wyemitowanych w ciągu dwóch lat 108 odcinków słuchowiska radiowego, w którym kwestie te były poruszane [Babbie 2003: 370]. Także w badaniach stosujących zogniskowany wywiad grupowych, w których nie dysponujemy co prawda danymi ilościowymi, możemy jednak zaobserwować bardzo wyraźne zmiany stanowisk pod wpływem udziału w dyskusji. Dobrym przykładem pokazującym takie oddziaływanie mogą być badania Ewy Nowickiej i jej zespołu nad tolerancją wobec innych etnicznie i kulturowo. Pod wpływem dyskusji, podczas której ukryty pomocnik moderatora

<sup>17</sup> Dla przypomnienia, w sondażu deliberatywnym<sup>®</sup> zorganizowanym w Poznaniu największa odnotowana różnica wynosiła 19 punktów procentowych.

<sup>18</sup> W roku 1994 blisko 72% badanych zadeklarowało stosowanie zasad związanych z profilaktyką dotyczącą HIV/AIDS, podczas gdy rok później – po wysłuchaniu serialu radiowego – takie deklaracje złożyło już 82% respondentów.

wygłaszał rasistowskie poglądy, norma tolerancji wobec innych została – jak to określają autorzy badania – „zawieszona”, dzięki czemu uczestnicy dyskusji zaczęli swobodnie prezentować nietolerancyjne poglądy [Nawrocki 1996: 91–93].

Chcemy podkreślić, że oprócz wyżej opisanych wątpliwości dotyczących sondażu deliberatywnego®, zauważamy także bezsporne zalety tej metody. Uważamy, że jedną z najważniejszych, jest próba włączenia obywateli w proces podejmowania decyzji. W trakcie deliberacji poznajemy nie tylko opinie ich uczestników, ale co dla nas szczególnie istotne, mamy możliwość poznania argumentów, które stoją za tymi opiniami. Współwystępowanie ze sobą postępowania zgodnego z metodologią badań ilościowych (dwa pomiary za pomocą kwestionariusza wywiadu lub ankiety) z postępowaniem jakościowym (zapis dyskusji grupowych) jest dodatkowym walorem tej metody, która pozwala w trakcie jednego badania zgromadzić tak różnorodne, a zarazem potwierdzające czy uzupełniające się dane. W „zwykłym” postępowaniu badawczym, jeśli badacz chciałby uzyskać pogłębienie opinii wypowiedzianych w sondażu – badanie jakościowe następowałoby po ilościowym i najprawdopodobniej obejmowałoby celowo dobranych respondentów.

Kolejnym ważnym aspektem tego typu podejścia jest próba zaktywizowania jednostek, przezwyciężenia obojętności obywateli. Oferując badanym ciekawe doświadczenie (możliwość spotkania z ekspertami, przebywania poza swoim naturalnym środowiskiem, w którym na co dzień wymieniają poglądy, poznanie nowych miejsc, nauczenie się nowych reguł dyskusji) można „zaszczepić w nich ducha obywatelskości”. Może dzięki takiej dyskusji chętniej będą głosowali w wyborach, wezmą udział w kolejnych konsultacjach społecznych, czy nawet wyjdą z własną inicjatywą, np. dotyczącą poprawy warunków ich życia. Ludzie, których pyta się o zdanie w sprawach istotnych dla miasta, kraju, Europy są bardziej dowartościowani i zaczynają czuć, iż mają wpływ na to, co się wokół nich dzieje<sup>19</sup>. Zdają sobie sprawę, że nie tylko eksperci czy politycy mają prawo prezentować swoje zdanie. Ważne są też nowe umiejętności związane z prowadzeniem dyskusji, dochodzeniem do kompromisu, wszak

[...] deliberacja promuje wzajemny szacunek, poprawia zrozumienie problemów, potrzeb i perspektyw drugiej strony, a w rezultacie kształtuje pozytywne nastawienie pomiędzy uczestnikami. Deliberatywne myślenie w kategoriach dobra publicznego buduje poczucie wspólnoty pomiędzy zaangażowanymi osobami. W jego wyniku może nastąpić zbliżenie lub dopasowanie rozbieżnych dotychczas stanowisk. Siły polityczne będą wówczas uzupełniane o tzw. kapitał społeczny w postaci zaufania społecznego, upowszechniania nastawienia na dialog i tolerancję, a także

<sup>19</sup> W badaniach prowadzonych przez CBOS procent respondentów, którzy odpowiadają twierdząco na pytanie o wpływ, jaki zwykli ludzie mają na sprawy w kraju nigdy nie przekracza 30% [Feliksiak 2009: 1]

zmniejszenia uprzedzeń. Wytwarza się wówczas więź pomiędzy członkami społeczeństwa oraz pomiędzy obywatelami a instytucjami politycznymi. Procesy te są szczególnie ważne dla scalania i konsolidowania nowoczesnych, heterogenicznych społeczności [Wesołowska 2010: 32].

Nagłośnienie przez media decyzji podjętych przez uczestników sondażu deliberatywnego® nakłada na decydentów obowiązek wzięcia ich pod uwagę. „Polityczna siła werdyktu nie jest jednak mała już choćby ze względu na jego medialny wydźwięk – decyzję bowiem powzięli przedstawiciele społeczeństwa obywatelskiego, którym starano się stworzyć najlepsze warunki do jej sformułowania” [Sroka 2009: 122]. Decyzje, którym towarzyszą argumenty przedstawione przez wielu dyskutujących, wydają się niezależne „[...] zarówno metodologicznie, jak i organizacyjnie” [Juhacz 2006: 52] od władz, które powinny je uwzględnić, a także prowadzą do podejmowania bardziej przemyślanych działań.

Podkreślić należy, iż przeprowadzenie deliberacji (w odróżnieniu od konsultacji społecznych czy wysłuchań publicznych) pozwala w wielu przypadkach na wypracowanie konsensusu w ważnych kwestiach i rozwiązywanie lub przynajmniej rozładowywanie konfliktów, jakie te kwestie wywołują. W efekcie deliberacji otrzymuje się nie tylko wachlarz potencjalnych rozwiązań, ale wskazuje się na to, do którego większość dyskutujących się skłania. Podkreśla się, że czasami w trakcie deliberacji, dzięki m.in. panującej tam atmosferze, pojawiają się rozwiązania, które bez niej nie mogłyby zaistnieć – „[...] w toku deliberacji mogą się ujawnić alternatywy przyszłych działań, których dostrzeżenie bywa zgoła niemożliwe w zwykłych okolicznościach decyzyjnych” [Sroka 2009: 128]. Nawet, jeżeli nie osiągniemy porozumienia, to mamy przynajmniej jasność, w czym się zgadzamy, a w jakich sprawach nie potrafimy wypracować porozumienia.

## BIBLIOGRAFIA

- B a b b i e E. [2003], *Badania społeczne w praktyce*, Warszawa: Wydawnictwo Naukowe PWN.
- B a u m a n Z. [2000], *Globalizacja*, Warszawa: PIW.
- B o g u s z e w s k i R. [2009], *Prywatyzacja – oceny, skojarzenia, oczekiwania i obawy*, Komunikat CBOS BS 133/2009.
- B o h m a n J. [1998], *The coming of age of deliberative democracy*, “British Journal of Political Science”, nr 5(1).
- B u k s i ń s k i T. [2001], *Dylematy demokracji deliberatywnej Johna Rawlsa i Jürgena Habermasa*, [w:] R. Marszałek, E. Nowak-Juchacz (red.), *Rozum jest wolny, wolność – rozumna*, Warszawa: Wydawnictwo IFiS PAN.
- B u k o w s k i A. [2011], *Sondaż deliberatywny*, [w:] S. Rudnicki (red.), *Nowe perspektywy. Nauki społeczne dla gospodarki*, Wyższa Szkoła Europejska im. ks. Józefa Tischnera, Kraków.
- C z y ż e w s k i M., D u n i n K., P i o t r o w s k i A. [2010], *Cudze problemy, czyli wstęp do sepologii*, [w:] M. Czyżewski, K. Dunin, A. Piotrowski (red.), *Cudze problemy. O ważności*

- tego, co nieważne. Analiza dyskursu publicznego w Polsce, Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Downs A. [1965], *Nonmarket Decision Making. A Theory of Bureaucracy*, "American Economic Review", nr 2.
- Dukaczewska-Nałęcz A. [1999], *Zogniskowane wywiady grupowe – jakościowa technika badawcza*, [w:] H. Domański, K. Lutyńska, W.A. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa: Wydawnictwo IFiS PAN.
- Elster J. [1989], *Deliberation and constitution making*, [w:] J. Elster, *Deliberative Democracy*, Cambridge: Cambridge University Press.
- Feliksiak M. [2009], *Opinie o funkcjonowaniu demokracji w Polsce*, Komunikat CBOS BS 20/2009.
- Feliksiak M. [2011], *Ocena działalności parlamentu, prezydenta i wybranych instytucji publicznych*, Komunikat CBOS BS 8/2011.
- Fishkin J. S. [1991], *Democracy and Deliberation: New Directions for Democratic Reform*, New Haven: Yale University Press.
- Głock Ch. Y. [1965], *Metoda panelowa*, [w:] S. Nowak (red.), *Metody badań socjologicznych. Wybór tekstów*, Warszawa: PWN.
- Gostkowski Z. [1966], *O założeniach i potrzebie badań nad technikami i procedurami badawczymi*, [w:] Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław: Ossolineum.
- Grzeszkiewicz-Raduńska K. [2009], *Respondenci niedostępni w badaniach sondażowych*, „Analizy i próby technik badawczych w socjologii”, t. XII, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Held D. [2010], *Modele demokracji*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Hipsz N. [2011], *Opinie o prawnej dopuszczalności i regulacji aborcji*, Komunikat CBOS BS 102/2011.
- Juchacz P.W. [2006], *Demokracja – Deliberacja – Partycypacja. Szkice z teorii demokracji ateńskiej i współczesnej*, „Pisma Filozoficzne”, t. XCIX, Poznań: Wydawnictwo Naukowe IF UAM.
- Koseła K., Sułek A. [2005], *Badania wzdłuż czasu*, [w:] *Encyklopedia socjologii. Suplement*, Warszawa: Oficyna Naukowa.
- Kubiak A. [2007], *Destrukcyjność procesu akulturacji badań sondażowych w Polsce*, „Przegląd Socjologiczny”, t. LVI/1.
- Nawrocki J. [1996], *Tolerancja z ograniczeniami. Raport z badań jakościowych*, [w:] E. Nowicka, J. Nawrocki (red.), *Inny – obcy – wróg. Swoi i obcy w świadomości młodzieży szkolnej i studenckiej*, Warszawa: Oficyna Naukowa,
- Nie chcemy, by Lech zarządził stadionem miejskim!*, [www.mm.poznan.pl](http://www.mm.poznan.pl) [dostęp 23.11.2009].
- Nowak S. (red.) [1965], *Metody badań socjologicznych. Wybór tekstów*, Warszawa: PWN.
- Przybylska A., Siu A. [2010], *Co Poznaniacy zrobią ze stadionem? Wyniki sondażu deliberatywnego® w Poznaniu*, <http://www.ps2012.pl/index.php/Konsultacje/Sondaz-deliberatywny-w-Poznaniu> [data dostępu 25.05.2012]
- Przybyłowska I. [1996], *Proces akulturacji sondaży opinii publicznej w Polsce*, „Przegląd Socjologiczny”, t. XLV.
- Reykowski J. [2007], *Program badań pt. „Psychologiczne uwarunkowania i konsekwencje deliberatywnego podejmowania decyzji w grupach politycznych”*. Załącznik [w:] J. Reykowski

- ski (red.), *Konflikt i porozumienie. Psychologiczne podstawy demokracji deliberatywnej*, Warszawa: Wydawnictwo SWPS Academica.
- Rozmowa z prof. Jamesem Fishkinem, socjologiem z Uniwersytetu Stanforda [2009], rozmawiała L. Włodek-Biernat, „Gazeta Wyborcza”, 23.11.
- Schwarz R.M. [2002], *The Skilled Facilitator: Practical Wisdom for Developing Effective Groups*, San Francisco: Jossey-Bass.
- Shapiro I. [2006], *Stan teorii demokracji*, Warszawa: Wydawnictwo Naukowe PWN.
- Sondaż deliberacyjny w Poznaniu. Zagospodarowanie Stadionu Miejskiego przy ul. Bułgarskiej po EURO 2012 [2009], <http://www.ps2012.pl/index.php/Konsultacje/Sondaz-deliberacyjny-w-Poznaniu> [data dostępu 25.05.2012].
- Sroka J. [2009], *Deliberacja i rządzenie wielopasmowe. Teoria i praktyka*, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Sułek A. [2001], *Funkcje badań opinii publicznej w ustroju demokratycznym. Przypadek Rzeczypospolitej Polskiej*, [w:] idem, *Sondaż Polski*, Warszawa: Wydawnictwo IFiS PAN.
- Szlachta B., Krzyżówek A. [2006], *Polityka deliberatywna: realistyczna utopia czy zbędny eksperyment?*, [w:] A. Mania, P. Laider (red.), *Amerykańska demokracja w XXI wieku*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Szymborski K. [2005], *Jarmark idei. Kiedy tłum ujawnia mądrość?*, „Niezbędnik Inteligenta”, dodatek do „Polityki”, <http://www.polityka.pl/spoleczenstwo/niezbednikinteligenta/12868,1,kiedy-tlum-ujawnia-madrosz.read> [data dostępu 25.05.2012].
- Walzer M. [2006], *Polityka i namiętność. O bardziej egalitarny liberalizm*, Warszawa: Muza SA.
- Wasilewski J. [2007], *Czy demokracja deliberatywna naprawi współczesną demokrację?*, [w:] J. Reykowski (red.), *Konflikt i porozumienie. Psychologiczne podstawy demokracji deliberatywnej*, Warszawa: Wydawnictwo SWPS Academica.
- Wesołowska E. [2010], *Deliberatywne rozwiązywanie konfliktów wartości. Wiele dróg do porozumienia*, Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego.
- Wnuk-Lipiński E. [2004], *Świat międzyepoki*, Kraków: Wydawnictwo Znak.

Anna Kubiak, Aneta Krzewińska

#### DELIBERATIVE POLLING® – A LIST OF PROBLEMS

##### Abstract

In this paper we presented deliberative polling® – the opinion research method created and patented by James S. Fishkin. In the first part of the article we focused on phases of conducting this type of research and presented the only one so far in Poland deliberative polling® which was organized in Poznań in 2009. The researchers were interested in the opinions of residents on the management of the stadium after EURO 2012. In the second part we analysed the research method from a variety of points of view, we highlighted some of the problems connected with organizing of this type of research, but we underline advantages of this research. The most important advantages included on: stimulating of citizens to act and contribute into the process of making decisions.

**Keywords:** deliberative polling®, deliberative democracy, deliberation

FRANCISZEK SZTABIŃSKI  
ORBS IFiS PAN  
TERESA ŻMIJEWSKA-JĘDRZEJCZYK  
ORBS IFiS PAN/MillwardBrown SMG/KRC

## ***MIXED MODE SURVEY DESIGN: PROBLEM EFEKTU TECHNIKI***

### **Streszczenie**

W sytuacji pogarszających się warunków realizacji badań surveyowych poszukuje się sposobów ograniczenia błędu pomiaru (*Total Survey Error*), związanego z nieuczestniczeniem wylosowanych jednostek w wywiadzie (*Non Response Error*). Jedno z rozwiązań stanowi procedura *Mixed Mode Survey Design*, która polega na uzyskiwaniu tych samych informacji od różnych osób, przy wykorzystaniu odmiennych technik badawczych, np. wywiad PAPI (*Paper and Pencil Interviewing*), CATI (*Computer Assisted Telephone Interviewing*), CAWI (*Computer Assisted Web Interviewing*). Jednakże wykorzystanie tej procedury może powodować wystąpienie efektu techniki (tzw. *Mode Effects*). W artykule analizujemy efekt techniki na płaszczyźnie: (i) sformułowania pytania; (ii) sposobu kontaktowania oraz (iii) sposobu komunikowania się z respondentem, poddając ocenie wskaźnik realizacji (tzw. *Response Rate*) oraz porównując odpowiedzi respondentów na te same pytania, uzyskane przy zastosowaniu różnych technik.

W analizach wykorzystano materiały zebrane w projektach realizowanych techniką PAPI (badanie European Social Survey, round 3) oraz CATI (badanie w ramach metodologicznego grantu ESS Infrastructure: i3.). Wyniki przeprowadzonych analiz potwierdzają istnienie poważnych różnic związanych z efektem techniki (w przypadku wywiadu PAPI i CATI). Stawia to pod znakiem zapytania ideę ich łączenia celem ograniczenia błędu braku odpowiedzi.

**Słowa kluczowe:** *Total Survey Error*, błąd braku odpowiedzi, efekt techniki, *Mixed Mode Survey Design*, Europejski Sondaż Społeczny

---

\* fsztabin@ifispan.waw.pl \*\* tzmijews@ifispan.waw.pl

## UWAGI WSTĘPNE

Z założenia badania surveyowe uznawane są za pomiar, a tym samym mają umożliwiać ściśle oszacowanie liczbowe badanych zjawisk i uogólnianie wniosków na całą populację. Aby było to możliwe, muszą zostać spełnione trzy podstawowe warunki: masowość (chodzi o liczebność próby do badania), losowy charakter próby oraz standaryzacja narzędzia, czynności badawczych i warunków realizacji badania. Każdy z tych elementów procesu badawczego jest jednak obarczony ryzykiem wystąpienia błędu. Powszechnie przyjętym obecnie sposobem oceny błędu w badaniach surveyowych jest koncepcja *Total Survey Error* (TSE) [Weisberg 2005; de Leeuw, Hox, Dillman 2008a: 1–17; Groves 1991]<sup>1</sup>. W klasycznym ujęciu, TSE obejmuje cztery rodzaje błędów: (i) błąd próby, (ii) pokrycia, (iii) błąd braku odpowiedzi oraz (iv) pomiaru. Błąd próby (tzw. *Sampling Error*) jest efektem tego, że badanie jest realizowane wśród wylosowanych osób, a więc na części, nie zaś na całej zbiorowości. Błąd pokrycia (tzw. *Coverage Error*) polega na tym, że operat losowania nie obejmuje wszystkich jednostek, wśród których chcemy zrealizować badanie, jest niekompletny bądź występują w nim „powtórzenia” (niektóre jednostki występują w nim więcej niż jeden raz). Z kolei błąd braku odpowiedzi (tzw. *Non-Response Error*) obejmuje dwie kategorie [Groves 1989; Grzeszkiewicz-Radulska 2009: 21–36]. Pierwsza z nich to nieuczestniczenie w badaniu wylosowanych jednostek, przede wszystkim ze względu na brak możliwości kontaktu z nimi oraz odmowy uczestnictwa w badaniu (tzw. *Unit Nonresponse*). Drugi rodzaj błędu braku odpowiedzi to brak odpowiedzi respondenta na dane, konkretne pytanie (tzw. *Item Nonresponse*). Ostatni rodzaj błędu to błąd pomiaru (tzw. *Measurement Error*), który jest „sumą” błędów związanych ze sposobem uzyskiwania danych (techniką) i sformułowaniem pytań (tzw. błąd narzędzia) oraz błędów związanych z osobą ankietera i osobą respondenta<sup>2</sup>.

<sup>1</sup> Koncepcja ta korzysta bezpośrednio z zaproponowanej przez Roberta Grovesa [1989] klasyfikacji błędów. Autor ten mówi o dwóch podstawowych rodzajach błędów: niezwiązanych z obserwacją oraz związanych z nią. Wśród tych pierwszych wyróżnia błąd: losowania próby, „pokrycia” oraz „braku odpowiedzi” (na poziomie całego wywiadu i poszczególnych pytań). Do drugiej kategorii błędów Groves zalicza błąd związany ze sposobem uzyskiwania danych (techniką); z pytaniem (określany także błędem narzędzia) lub z osobą ankietera oraz osobą respondenta.

<sup>2</sup> Rozwinięcie klasycznej koncepcji TSE przedstawiła m.in. Edith de Leeuw [de Leeuw et al. 2008a], wprowadzając piąty element – błąd specyfikacji (tzw. *specification error*) występujący w sytuacji, gdy pytanie mierzy różny od założonego koncept, np. gdy obserwujemy niską trafność konstrukcyjną (tzw. *construct validity*). Nie wspomina ona jednak o błędach opracowania (tzw. *processing errors*), a więc błędach: kodowania, wprowadzania, danych, edycji danych oraz



Obecnie przyjmuje się, że błąd braku odpowiedzi na poziomie całego wywiadu jest jednym z najpowszechniejszych i to właśnie on decyduje o rozmiarze TSE. Praktycznie od połowy lat osiemdziesiątych obserwuje się ciągły spadek tzw. odsetka realizacji (*Response Rate*)<sup>3</sup>. Tendencja ta występuje na całym świecie, choć jej skala jest różna w zależności od techniki i tematyki badania [Goyder 1987; de Leeuw, de Heer 2002: 41–54; Groves et al. 2004: 185–187; Stoop et al. 2010]. W Holandii odsetek realizacji w badaniach czytelnictwa spadł poniżej 50% w 1998 r., a w 2000 r. nie osiągnął nawet 40% [Bassler A., Tchaoussoglou C. 2001]. W Polsce, w projekcie Polski Generalny Sondaż Społeczny [Cichomski et al. 2009] odsetek realizacji zmniejszył się na przestrzeni ostatnich 16 lat (1992–2008) o 30 punktów procentowych (z 82 do 52%), co oznacza, że w 2008 r. nie zrealizowano 48% wylosowanej próby. Edith de Leeuw i Wim de Heer [de Leeuw, de Heer 2002], analizując badania realizowane w latach 1980–1989 w 16 krajach (w tym Stany Zjednoczone i Kanada), sformułowali prognozę, iż średni wzrost odsetka braku kontaktu (*Non-Contacts*) będzie postępował w tempie 0,2% na rok, zaś wskaźnik odmów (*Refusals*) – w tempie 0,3%. Nie należy się dziwić, że w 2005 r. ukazała się książka Ineke Stoop zatytułowana *The Hunt for the Last Respondent: Nonresponse in Sample Surveys* [Stoop 2005].

O wadze błędu braku odpowiedzi na poziomie wywiadu świadczy także to, że nie ma on charakteru losowego, jest błędem systematycznym i powoduje tendencyjność wyników: osoby niebiorące udziału w badaniu różnią się w sposób istotny – ze względu na cechy społeczno-demograficzne, psychologiczne oraz opinie – od osób, które w nim uczestniczyły [Sztabiński P.B., Sztabiński F., Przybysz 2007: 26–54].

Czy zatem w sytuacji narastającego problemu błędu braku odpowiedzi (*Non-Response Error*) istnieją sposoby jego ograniczenia? Zasadniczo tak, aczkolwiek winno się raczej mówić o sposobach „zapobiegania” powstawaniu tego błędu (*ex ante*) oraz sposobach jego „naprawiania” (*ex post*), czyli o dwu różnych, komplementarnych podejściach. W ramach pierwszego z nich wymienić należy wszystkie działania z zakresu Quality Control lub Quality Assurance, mające na celu optymalizację przebiegu procesu badawczego z punktu widzenia jakości uzyskiwanych danych (chodzi o właściwe przygotowanie narzędzia badawczego, szkolenie i monitoring ankierów itd. [Sztabiński F. 2011: 19–35]. W ramach drugiego podejścia wskazuje się przede wszystkim ważenie poststratyfikacyjne

uzupełnienia braków danych (tzw. *imputation errors*). [Biemer, Caspar 1994: 307–326; Weisberg 2005].

<sup>3</sup> Mówimy tutaj oczywiście o badaniach, które są realizowane na próbach bez zastępowania (tzn. realizowanych „do wyczerpania”).

(uwzględniające rozkład cech respondentów w próbie populacji) oraz ważenie w oparciu o tzw. wskaźnik dostępności respondenta (*propensity score*), który określa prawdopodobieństwo udziału w badaniu w zależności od jego cech, nie tylko demograficznych [Stoop et al. 2010].

Zupełnie inne od wskazanych powyżej, aczkolwiek mieszczące się w ramach procedur *ex ante*, jest tzw. *Mixed Mode Survey Design*, a więc wykorzystanie różnych technik zbierania danych w ramach jednego projektu badawczego. Wiadomo, że właściwy dobór techniki badania może przynajmniej w pewnym zakresie ograniczyć błąd braku odpowiedzi. Chodzi tutaj o problem tzw. adekwatności techniki badania [Gostkowski 1965, 1966; Słomczyński 1968; Sztabiński F. 1997, 2003: 147–175]. W uproszczeniu sprowadza się on do tego, że różni ludzie preferują różne sposoby kontaktowania się i komunikowania. Dodatkowo, niezależnie od preferencji, nie ze wszystkimi można się skontaktować w dany sposób (np. młodzi ludzie kształcący się „poza domem” są trudniej dostępni pod adresem zameldowania, reprezentanci określonych grup zawodowych – głównie profesjonaliści i drobni przedsiębiorcy – są łatwiej dostępni „pod telefonem”).

Jeśli błąd braku odpowiedzi można ograniczyć, łącząc różne techniki zbierania danych w ramach jednego badania, to rodzi się pytanie, w jakim stopniu taki „cocktail” technik może jednocześnie zwiększać błąd pomiaru wynikający z efektu techniki (*Mode Effects*). Jest rzeczą oczywistą, że poszczególne techniki surveyowe mają sobie właściwe nie tylko zalety, ale i ograniczenia. Wiadomo, iż te same osoby, w zależności od formy kontaktu (osobisty lub pośredni, w tym telefoniczny, przez Internet itp.) oraz kanału komunikowania (komunikacja ustna lub pisemna) mogą w różny sposób reagować na ten sam bodziec, a więc udzielać różnych odpowiedzi na to samo pytanie<sup>4</sup>. Skoro tak, czy można je kwantyfikować, a następnie uogólniać na całą populację? Raczej nie, co przemawia za tym, aby nie „mieszać” technik. Na badanie należy jednak spojrzeć całościowo, z punktu widzenia „strat” i „korzyści”. *Przede wszystkim trzeba ocenić, czy wprowadzenie różnych technik (tzw. Mixed Mode) zmniejszy błąd braku odpowiedzi (Nonresponse Error), czy też zwiększy błąd pomiaru (Measurement Error) przy danych kosztach. Jest to swoisty trade-off: czy lepiej nie „mieszać” technik i uzyskać odpowiedzi od połowy lub mniejszego odsetka wylosowanych osób, czy też – mając świadomość błędu pomiaru związanego z efektem techniki – lepiej je łączyć i uzyskać odsetek realizacji na poziomie 70% lub powyżej<sup>5</sup>.*

<sup>4</sup> Wiąże się to przede wszystkim z dystansem „sytuacyjnym” i „czasowym”, charakterystycznym dla komunikacji pisemnej [Sztabiński F. 1997: 41–47].

<sup>5</sup> Odsetek taki stanowi minimum w badaniach europejskich [European Social Survey 2011].

W niniejszym artykule podejmujemy problem efektu techniki (*Mode Effects*) i jej wpływu na odpowiedzi respondentów, a tym samym rozmiar błędu pomiaru w ramach TSE. Ponieważ efekt techniki oznacza odmiennosc: (i) sformułowania pytania, (ii) sposobu kontaktowania oraz (iii) sposobu komunikowania się z respondentem, przedstawimy przede wszystkim zakres i rozmiar owych różnic. W analizach wykorzystane zostaną materiały zebrane w projekcie zrealizowanym techniką PAPI – ESS r. 3 oraz techniką CATI – ESS r. 3 zastosowaną w projekcie metodologicznym ESS Infrastructure: i3. Najpierw jednak doprecyzujemy termin *Mixed Mode Survey Design* oraz przedstawimy ogólne wnioski z badań nad różnymi technikami.

### **MIXED MODE SURVEY DESIGN**

*Mixed Mode Survey Design* (*Mixed Mode Data Collecting*, *Mixed Mode Survey*, *Mixed Mode Design* itp., w skrócie *Mixed Mode*) jest procedurą badania, w której dopuszcza się łączenie nie tylko różnych technik wywiadu (np. PAPI i CATI), lecz także technik wywiadu z technikami ankiety (np. CATI, CAWI lub Ankieta Pocztaowa) [de Leeuw et al., 2008b: 299–316]. Z jednej strony, idea łączenia technik nie jest niczym nowym (np. zastosowanie ankiety wypełnianej pod kontrolą ankietera w trakcie lub po wywiadzie). Z drugiej jednak jest ona zupełnie nowa, ponieważ zakłada uzyskiwanie danych – w ramach tego samego badania – różnymi technikami.

Czy jednak w obu przypadkach mówimy o tym samym? Raczej nie. W związku z tym celowe wydaje się doprecyzowanie terminu *Mixed Mode*. O zróżnicowaniu stosowanych technik można mówić na etapie kontaktowania się z respondentem oraz zbierania danych. W pierwszym przypadku wykorzystujemy różne sposoby, kanały komunikowania się (np. pisemny – listy zapowiednie, ustny – telefony do respondentów). Nie nazwiemy tego *Mixed Mode*. W drugim przypadku (etap zbierania danych) w wielu projektach przyjęty schemat badania zakłada wykorzystanie różnych technik zbierania danych. Jak wspomniano, w trakcie lub po wywiadzie (PAPI/CAPI) możemy wręczyć respondentowi ankietę do samodzielnego wypełnienia<sup>6</sup>. Tego badania również nie nazwiemy *Mixed Mode*, bowiem w przypadku każdego respondenta stosujemy ten sam schemat, a zatem możemy co najwyżej mówić o technice zasadniczej i pomocniczej (dodatkowej).

---

<sup>6</sup> Taki schemat badania stosuje się w Polskim Generalnym Sondażu Społecznym (PGSS). Po zrealizowaniu wywiadu ankieter wręcza respondentowi ankietę ISSP do samodzielnego wypełnienia (w obecności ankietera).

Don Dillman [2000: 219–222] mówi o pięciu sytuacjach, które można określić mianem *Mixed Mode* [także: de Leeuw, Hox, Dillman, 2008b: 299–316]. Są one następujące:

– zbieranie tych samych informacji, od różnych osób przy zastosowaniu różnych technik; jest to najpowszechniej stosowana procedura i podstawowym jej celem jest ograniczenie błędu braku odpowiedzi, przy danych lub zredukowanych kosztach;

– zbieranie danych „panelowych” różnymi technikami od tych samych osób w kolejnych falach badania;

– zbieranie różnych informacji różnymi technikami od tych samych osób wylosowanych w trakcie „jednorazowego” badania (wspomniana sytuacja, kiedy w trakcie lub po wywiadzie zasadniczym pozostawia się respondentowi ankietę do wypełnienia);

– zbieranie tych samych informacji różnymi technikami w różnych populacjach (w projektach międzykrajowych, w poszczególnych krajach badanie może być realizowane różnymi technikami. Europejski Sondaż Społeczny w niektórych krajach realizuje się techniką PAPI, a w innych CAPI [European Social Survey 2011].

– wykorzystanie różnych technik „wspomagających” realizację badania prowadzonego inną techniką – mogą to być listy zapowiednie w przypadku badań *face-to-face* czy monity telefoniczne w przypadku badań realizowanych techniką ankiety pocztowej.

Przedstawione przez Dillmana rozumienie *Mixed Mode* należy uznać za zbyt szerokie, ponieważ większość badań należałoby zaliczyć do tej kategorii. Jeśli przed badaniem realizowanym techniką PAPI wysyłamy do respondentów list zapowiedni, to należałoby uznać, że jest to badanie *Mixed Mode*<sup>7</sup>. Podobne, węższe rozumienie tego terminu przyjmują Edith de Leeuw [2009], Peter Lynn [2009] oraz Peter Martin [2010], mając na myśli wyłącznie pierwszą z wyróżnionych sytuacji, a więc zbieranie tych samych informacji przy zastosowaniu różnych technik od różnych jednostek stanowiących elementy danej próby. W pozostałych przypadkach należałoby mówić o *Multi(ple) Mode Design*.

Jeśli akceptujemy ideę *Mixed Mode*, rozumianą we wskazany sposób, musimy odpowiedzieć sobie na dwa podstawowe pytania: w jaki sposób łączyć techniki i jakie („How to Mix?”, „What to Mix?”). Zasadniczo istnieją *dwa sposoby implementacji Mixed Mode* [de Leeuw, Hox, Dillman 2008a: 306–307, 310–312; Martin

<sup>7</sup> Warto tutaj zaznaczyć, iż w późniejszej pracy autor [Dillman et al. 2009] mówi już tylko o czterech typach *Mixed Mode* (pomija typ iv.).

2010]. Jeden z nich to sekwencyjne stosowanie różnych technik (tzw. *sequential*), które polega na tym, że określamy jedną podstawową (preferowaną) technikę oraz dodatkowe, uzupełniające, które będą stosowane w dalszej kolejności. Przykładowo, jeśli podstawową techniką jest CATI, to do grupy respondentów, z którymi nie udało się zrealizować wywiadu telefonicznie, możemy wysłać ankieterów, aby zrealizowali z nimi wywiad *face-to-face* (techniką PAPI lub CAPI). Celem sekwencyjnego *Mixed Mode* jest oczywiście redukcja błędu braku odpowiedzi (*unit nonresponse*), ale także kosztów badania. Zazwyczaj jako technikę podstawową wybiera się najtańszą, np. ankietę Web-ową lub tradycyjną ankietę pocztową. Kolejne techniki są relatywnie droższe w zastosowaniu (np. wywiad CATI, CAPI czy PAPI). Drugi rodzaj *Mixed Mode* polega na równoczesnym (tzw. *concurrent*) stosowaniu różnych technik zbierania informacji. Nie określa się wówczas techniki podstawowej, lecz traktuje wszystkie jak ekwiwalentne i wykorzystuje się je równocześnie, w tym samym czasie. Możemy wysłać do wylosowanych osób list z informacją o badaniu i propozycją wypełnienia ankiety „papierowej” (ankieta pocztowa), e-mailowej lub Web-owej. W zależności od preferencji wysyłamy respondentowi właściwą wersję ankiety lub też proponujemy rozmowę telefoniczną lub osobistą. Celem równoczesnego *Mixed Mode Design* jest, poza redukcją błędu braku odpowiedzi, przede wszystkim redukcja błędu pokrycia (np. osoby niemające dostępu do Internetu mogą wypełnić ankietę „papierową”), ale także minimalizacja kosztów.

Odpowiedź na pytanie o sposób administrowania poszczególnych technik (w sposób sekwencyjny czy równoczesny) zależy przede wszystkim od tego, jakiego rodzaju błąd chcemy minimalizować przy danych kosztach. Zgodnie z koncepcją *Total Design Method* (TDM) Dillmana [1978; 2000], opartą na teorii wymiany społecznej, model równoczesny *Mixed Mode* winien być optymalny i bardziej efektywny niż sekwencyjny. Daje respondentowi poczucie podmiotowości, „możliwość wyboru formy kontaktu”, stanowi wyraz „dyspozycyjności” badacza wobec respondenta, przez co „dowartościowuje” go, a tym samym stanowi element szeroko rozumianego systemu nagród. Jednocześnie model równoczesny eliminuje negatywne implikacje związane z poczuciem „konieczności subordynacji”, minimalizując tym samym ponoszone przez respondenta „koszty”. Praktyka wskazuje, że model sekwencyjny jest bardziej efektywny, przynajmniej w zakresie kosztów<sup>8</sup>.

<sup>8</sup> Jaki więc model stosować? Wydaje się, iż zależy to przede wszystkim od badanej populacji. W badaniach *Mixed Mode* na populacji generalnej prawdopodobnie lepiej sprawdza się schemat sekwencyjny, natomiast w badaniach specyficznych kategorii respondentów chyba schemat równoczesny.

Właściwe stosowanie *Mixed Mode* z pewnością może redukować błąd braku odpowiedzi (na poziomie wywiadu), błąd pokrycia oraz koszty badania, komplikuje jednak schemat i sposób doboru próby (co może wpływać na wielkość błędu próby) oraz stwarza trudności, od strony logistycznej, w realizacji terenowej badania. Jeśli chodzi o błąd pomiaru, to zawsze należy liczyć się z efektem techniki (*Mode Effects*). I wreszcie koszty: zależą od tego, „co mieszamy i w jaki sposób”. W porównaniu z wywiadem PAPI (traktowanym jako technika podstawowa, *Single Mode*) zastosowanie sekwencyjne lub równoczesne *Mixed Mode*, w jakiegokolwiek konfiguracji, obniża koszty badania. Natomiast zastosowanie schematu sekwencyjnego, niezależnie od techniki podstawowej, podnosi koszty badania (rozpoczyna się od techniki „najtańszej”). Podsumowując, *Mixed Mode* odpowiednio stosowany powinien prowadzić do redukcji TSE, a w wielu przypadkach także kosztów.

### EFEKT TECHNIKI (*MODE EFFECTS*)

Ogólnie biorąc, efekt techniki polega na tym, że zastosowanie różnych sposobów zbierania materiałów (technik badawczych) może pociągać za sobą ryzyko uzyskania odmiennych, różnych danych. Ponieważ wcześniej wspominaliśmy, iż wykorzystanie *Mixed Mode* stanowi zawsze swoisty *trade-off* między redukcją błędu braku odpowiedzi a efektem techniki, warto w tym miejscu przypomnieć podstawowe charakterystyki niektórych technik badawczych, aby mieć świadomość, „co zyskujemy, a co tracimy”. Ograniczymy się do podstawowych technik *opartych na komunikacji ustnej wspomaganych lub niewspomaganych komputerowo* (CAPI, CATI, PAPI) oraz *technik opartych na komunikacji pisemnej, także wspomaganych lub niewspomaganych komputerowo* (CASI – *Computer Assisted Self Interviewing* oraz PSAQ – *Paper Self Administered Questionnaires*) [Saris: 1991]<sup>9</sup>.

<sup>9</sup> Ogólne rozważania na temat zalet i ograniczeń poszczególnych technik badawczych prowadzone były niemal „od zawsze”. W 1950 r. opublikowana została jedna z pierwszych prac poświęconych technikom surveyowym, oparta na ponad dwudziestoletnich doświadczeniach w dziedzinie badań tego typu [Parten, 1950]. Refleksja ta kontynuowana jest do dzisiaj, a problem określenia zalet i ograniczeń poszczególnych procedur badawczych rodzi się zawsze w związku z pojawieniem się nowych technik, np. CAPI, CATI czy CAWI [Bowling 2005: 281–291; de Leeuw et al. 2008a]. Na gruncie polskim tego typu rozważania prowadzone były m.in. przez J. Lutyńskiego [1968; 1972], F. Sztabińskiego [1997; 2009], P.B. Sztabińskiego [1999a–c; 2001: 65–89], F. Sztabińskiego i P.B. Sztabińskiego [1997: 73–92], T. Żmijewską-Jędrzejczyk [2004].

Zacznijmy od wywiadu kwestionariuszowego. Wybierając wywiad PAPI, który ma bardzo szeroki zakres zastosowania, mamy pewność, że w badaniu uczestniczy właściwa osoba, zapewniona jest standaryzacja sytuacji badania (chodzi o miejsce i warunki wywiadu). Dodatkowo, w przypadku pytań trudnych, niejasnych, istnieje możliwość pomocy i dodatkowych wyjaśnień ze strony ankietera (tym samym minimalizuje się ryzyko zbierania artefaktów). Zakłada się też, iż dzięki obecności ankietera zapewniona zostanie poprawność wypełnienia kwestionariusza. Jednocześnie ta sama technika ma określone ograniczenia. Przede wszystkim istnieje ryzyko wpływu ankietera na odpowiedzi respondenta [Lutyńska 1978, Sztabiński P.B. 1997] oraz, ze względu na ograniczone poczucie anonimowości badanych, ryzyko udzielania społecznie akceptowanych odpowiedzi. Jest to technika bardzo kosztowna, a długość okresu realizacji badania jest niemal wprost proporcjonalna do wielkości próby [Sztabiński F. 1997: 48–52].

Wywiad osobisty ze wspomaganiem komputerowym (CAPI), oprócz wspomnianych zalet wywiadu PAPI, ma jeszcze kilka dodatkowych. Przede wszystkim zastosowanie komputera stwarza możliwości wykorzystania stymulacyjnych materiałów wizualnych (np. przy testowaniu reklam), co w konsekwencji osłabia efekt znużenia respondenta. Umożliwia rotację kafeterii zawartej w pytaniu czy nawet całych bloków pytań, co niweluje efekt porządku. Technika CAPI daje ankieterowi poczucie profesjonalizmu i w znacznym stopniu standaryzuje jego zachowania poprzez ograniczenie błędów w prowadzeniu wywiadu, a także rozszerza zakres możliwej kontroli pracy ankieterów (chodzi m.in. o automatyczny zapis tzw. „międzyczasów” dla poszczególnych sekcji wywiadu). Co więcej, wykorzystanie komputera znacząco wpływa na obniżenie kosztów realizacji badania (eliminacja kosztów związanych z drukiem kwestionariuszy, zdjęć itd.) oraz skrócenie okresu realizacji badania (pominięcie fazy transportu wypełnionych kwestionariuszy ankiet i wprowadzania danych). Niemniej technika ta, obok zalet, ma także pewne ograniczenia. Oprócz analogicznych, jak dla wywiadu PAPI, istnieją jeszcze dwa, z których pierwsze może rzutować na wartość uzyskiwanych danych. Chodzi o problem redukcji wymiaru społecznego interakcji (koncentracja ankieterów przede wszystkim na komputerze i monitorze, nie zaś na interlokutorze) oraz poczucie respondentów zupełnego „prześwietlenia” i deprywatywacji (mówi się nawet o efekcie tzw. *Big Brother*) [Sztabiński P.B. 2001: 65–89]. Drugie ograniczenie ma charakter techniczny. Chodzi o takie sprawy, jak np. problemy z uruchomieniem czy podłączeniem komputera, błędy w skrypcie. Nie sposób nie wspomnieć także o poczuciu zagrożenia bezpieczeństwa osobistego ankieterów.

Kolejna technika – wywiad CATI – ma wiele analogicznych zalet jak wywiady PAPI i CAPI: bardzo szeroki zakres zastosowania, pewność, że w badaniu uczestniczy właściwa osoba, standaryzacja sytuacji badania (miejsca, warunków wywiadu, ograniczony zostaje czynny wpływ osób trzecich) oraz standaryzacja zachowań ankieterskich (poprzez bieżący monitoring zostaje ograniczone prawdopodobieństwo wystąpienia błędów ankieterskich, niezwiązanych z zadawaniem pytań, te „kontroluje” program). Istnieje także możliwość pomocy ze strony ankietera w przypadku pytań trudnych, niejasnych, a ze względu na większe poczucie anonimowości występuje słabsza (w porównaniu z innymi technikami) skłonność do udzielania społecznie akceptowanych odpowiedzi. Ze względu na charakter rozmowy telefonicznej (szybkie tempo) uzyskuje się „szybkie”, skrótowe i schematyczne, a w konsekwencji mało przemyślane i niepogłębione odpowiedzi (np. respondent przerywa ankieterowi odczytywanie kafeterii, gdy usłyszy pierwszą „satysfakcjonującą” go odpowiedź). Dodatkowo, ze względu na specyfikę kontaktu, występuje silniejsza, w porównaniu z wywiadem PAPI/CAPI, skłonność do „zgadzania się” (tzw. „potakiwania”) oraz wyboru skrajnych punktów skali, a przy skalach liczbowych 100-punktowych – częstszy wybór „pełnych” wartości. Wreszcie jedno z podstawowych ograniczeń wywiadu CATI: nie mamy możliwości wykorzystywania wizualnych materiałów stymulujących (przede wszystkim Kart Respondenta), aczkolwiek istnieje możliwość odtworzenia materiału dźwiękowego. Sposób kontaktowania się narzuca także pewne ograniczenia czasowe: wywiad nie może być zbyt długi, gdyż istnieje ryzyko przerwania rozmowy przez respondenta). W rezultacie techniki oparte na tym samym kanale komunikowania (komunikacja ustna), ale odmiennym sposobie kontaktowania (kontakt naoczny vs kontakt telefoniczny) mogą przynosić zupełnie różne wyniki. Przykładowo, dla pytań otwartych o znajomość firm ubezpieczeniowych różnice rozkładów odpowiedzi w badaniu CATI i PAPI sięgają 8–9 punktów procentowych (odsetki są niższe w CATI). Z kolei w przypadku pytań zamkniętych różnice rozkładów sięgają 20–30 punktów procentowych. W tym przypadku odsetek „znajomości” firm jest wyższy w CATI [Sztabiński P.B. 1999a]. Technika ta zapewnia także krótki czas realizacji badania i jest najtańszą techniką opartą na komunikacji ustnej. Jednocześnie stwarza ryzyko wypaczenia próby ze względu na sposób konstruowania próby do badania. Chodzi tutaj przede wszystkim o błąd pokrycia, wynikający z nierównomiernego wyposażenia gospodarstw domowych w telefony<sup>10</sup>.

<sup>10</sup> Obecnie, w związku ze wzrostem wyposażenia gospodarstw domowych w telefony komórkowe, obniża się poziom wyposażenia w telefony stacjonarne. Ponieważ do tej pory nie wypracowano



Jakie są zalety i ograniczenia ankiet samodzielnie wypełnianych przez respondentów w porównaniu z różnymi formami wywiadu? Pierwszą i podstawową jest to, że uzyskany materiał badawczy, w przeciwieństwie do technik wywiadu PAPI, CAPI i CATI, jest tożsamy ze źródłem informacji. Chodzi o to, że – ze względu na brak pośrednictwa ankietera – informacje, których udzielił respondent (sam zapisał) są identyczne z tymi, których chciał udzielić (w wywiadzie są „przefiltrowane” przez ankietera). Inne zalety ankiet to bardzo szeroki zakres zastosowania, poczucie anonimowości badanych oraz niski koszt realizacji. Jednakże, tak jak w przypadku innych technik, ankiety samodzielnie wypełniane przez respondenta mają pewne ograniczenia. Należy tutaj wymienić choćby to, iż nigdy nie ma pewności, kto faktycznie wypełnia ankietę (poza przypadkami ankiety pod kontrolą oraz audytoryjnej). Co więcej, respondenci zazwyczaj zapoznają się z treścią pytań ankiety przed jej wypełnieniem i mogą konsultować swoje odpowiedzi z innymi osobami (tzw. konsultacja „wewnątrzdomowa”), a w efekcie ich odpowiedzi są zakotwiczone społecznie, nie są zaś własne, indywidualne<sup>11</sup>. Inne ograniczenia sprowadzają się do znacznej pracochłonności jej przygotowania, dość długiego okresu realizacji badania (z monitami ok. 5–7 tygodni, przy czym długość okresu realizacji nie jest praktycznie zależna od wielkości próby), nieco niższego odsetka realizacji w porównaniu z wywiadem [Dillman 1978; Sztabiński F. 1997] oraz tego, że część ankiet wraca niewypełniona lub wypełniona niekompletnie. W efekcie wszystkich tych odmienności odpowiedzi uzyskiwane na te same pytania przynoszą inne rezultaty niż wywiad [Sztabiński F. 1997]. Dla pytań o opinie różnice rozkładów w ankiecie pocztowej i wywiadzie PAPI sięgają 17 punktów procentowych. W przypadku pytań faktualnych różnice zasadniczo nie występują.

Techniki ankiety ze wspomaganie komputerowym, a więc przede wszystkim CAWI. Łączą one wszystkie zalety technik ankiety i technik wspomaganych komputerowo. Oznacza to szybkość realizacji badania, niskie koszty przygotowania narzędzia i realizacji oraz możliwość wielokrotnego monitowania. Podstawowym ograniczeniem jest jednak relatywnie niski i bardzo zróżnicowany geograficznie poziom wyposażenia gospodarstw domowych w Internet (tzw. poziom penetracji, który w Polsce wynosi ok. 50%) i związany z tym błąd pokrycia [Dillman 2000; Sztabiński F. 2009; Żmijewska-Jędrzejczyk 2004].

---

efektywnej metodologii „łączenia” prób telefonów stacjonarnych i komórkowych, realizacja badań CATI bazuje na próbach telefonów stacjonarnych [Haeder et al. 2010: 3–9].

<sup>11</sup> Ostatnia cecha ankiety – społeczne zakotwiczenie opinii – nie dotyczy oczywiście ankiety pod kontrolą oraz audytoryjnej. Dyskusyjne jest to, czy owo „zakotwiczenie” społeczne jest ograniczeniem, czy też zaletą ankiet.

Znajomość zalet, ograniczeń i specyfiki poszczególnych technik jest niezbędna na etapie conceptualizacji badania, kiedy podejmujemy decyzję dotyczącą wyboru technik, które zamierzamy łączyć. Generalnie wydaje się, że należy dążyć do łączenia technik maksymalnie do siebie zbliżonych ze względu na sposób dotarcia i komunikowania, a unikać łączenia technik całkowicie odmiennych. Jednocześnie musimy uwzględnić problematykę badania, charakter badanej zbiorowości (chodzi tutaj o wspomniany problem społecznej i psychologicznej adekwatności techniki) oraz koszty badania. Wybierając sposób „mieszania technik”, musimy przede wszystkim odwoływać się do własnej wiedzy, doświadczenia badawczego oraz intuicji badawczej.

Ponieważ, jak wspomniano, efekt techniki wiąże się z odmiennością sformułowania pytań, odmiennością sposobu kontaktowania i komunikowania się z respondentem, rozważymy, jakie są konsekwencje owych odmienności i możliwości ograniczenia ich skutków. Przedtem jednak opiszemy projekt, którego rezultaty będziemy prezentować.

## OPIS PROJEKTU

Dane, na podstawie których będziemy prowadzić dalsze rozważania dotyczące efektu techniki zostały zebrane w ramach projektu JRA1 „Strategie zbierania danych: możliwości łączenia różnych technik zbierania danych w kolejnych rundach ESS”, który był częścią metodologicznego grantu ESS Infrastructure: i3. Projekt ten miał na celu m.in. podjęcie próby stworzenia narzędzia „odpornego” na efekt techniki oraz ocenę błędu braku odpowiedzi i błędu pomiaru związanego z daną techniką. Chodziło zatem o porównanie poszczególnych technik na dwóch płaszczyznach – realizacyjnej (jaki jest faktyczny *response rate* przy zastosowaniu różnych technik, przy zachowaniu porównywalności prób i narzędzi) i merytorycznej (w jakim stopniu dane uzyskane różnymi technikami są zbieżne).

Mimo że projekt JRA1 był realizowany od roku 2006, już wcześniej, bo w roku 2003 we współpracy z Gallup Europe przeprowadzone zostały (na Węgrzech, a w 2005 r. na Węgrzech i w Portugalii) wstępne badania metodologiczne dotyczące *Mixed Mode*. Na ich podstawie można stwierdzić, iż techniką najbardziej różniącą się od technik wywiadu osobistego (PAPI i CAPI) jest CATI, pomijając ankietę pocztową [por. Sztabiński P.B. 1999b: 51–66], jednocześnie żadna pojedyncza technika nie może być konkurencyjna wobec wywiadu PAPI [Eva, Widdop 2007]. Stąd też w niniejszym projekcie podjęto próbę udzielenia odpowiedzi na pytania w odniesieniu do techniki PAPI, CAPI oraz CATI.

Badanie ESS-CATI przebiegało wg schematu eksperymentalnego opracowanego przez Caroline Roberts i Petera Lynna, przy współpracy Gillian Eva. W projekcie porównano trzy techniki: CATI, CAPI i PAPI. Było to możliwe, ponieważ w projekcie uczestniczyło 5 krajów: Cypr, Niemcy, Polska, Szwajcaria oraz Węgry, z których niektóre realizowały badanie zasadnicze ESS r. 3 techniką PAPI, inne zaś CAPI. Porównanie dotyczyło ESS-CAPI i ESS-CATI (Niemcy, Szwajcaria) oraz ESS-PAPI i ESS-CATI (Cypr, Polska, Węgry).

Procedura badania ESS-CATI zakładała realizację trzech różnych wersji kwestionariusza ESS r. 3: A, B i C, które systematycznie alokowano w trakcie realizacji do wylosowanej próby. W grupie A, która stanowiła 40% próby (400 przypadków) realizowana była pełna wersja wywiadu ESS (czas trwania szacowany na ok. 60 min.); w grupie B, która stanowiła 20% próby (200 przypadków) realizowana była skrócona wersja wywiadu (czas trwania szacowany na ok. 45 min.); w grupie C, która stanowiła 40% próby (400 przypadków) realizowana była także pełna wersja wywiadu ESS, ale podzielona na dwie części (z każdą osobą realizowane były dwa wywiady po ok. 30 min. (zob. Aneks: Schemat badania ESS-CATI r. 3).

W Polsce badanie ESS-CATI przeprowadzał Instytut MillwardBrown SMG/KRC od grudnia 2006 do lutego 2007, pod kierownictwem i przy nadzorze metodologicznym IFiS PAN. Poszczególne fazy przygotowania i realizacji badania były następujące:

1. Przygotowanie „papierowych” wersji kwestionariuszy ESS-CATI oraz napisanie i testowanie skryptów trwało relatywnie długo (około 1 miesiąca). Po ostatecznym sprawdzeniu skryptu testowano go w terenie (przeprowadzono ponad 140 wywiadów pilotażowych).

2. Opracowanie założeń doboru próby. Próba do badania ESS-CATI z założenia miała być porównywalna z próbą w badaniu zasadniczym (ESS-PAPI). Ponieważ było to niemożliwe (badanie zasadnicze realizowano na próbie losowanej z operatu PESEL, w którym brak informacji o numerach telefonów), podjęto decyzję o doborze próby metodą List-Assisted RDD (Random Digit Dialing), a więc losowania końcówek numeru do istniejących i sprawdzonych prefiksów. Sposób warstwowania próby był identyczny jak w przypadku badania zasadniczego. Na etapie losowania respondenta zastosowano procedurę doboru Kisha. Badanie realizowano na 1000-elementowej próbie dorosłej ludności Polski w wieku 15 i więcej lat (bez górnej granicy), zamieszkującej w indywidualnych gospodarstwach domowych. Była to próba realizowana „do wyczerpania” (zastępowano tylko numery spoza próby – telefony nieistniejące oraz należące do

instytucji). Poszczególne wersje kwestionariusza ESS-CATI (A, B i C) były systematycznie alokowane do próby.

3. Szkolenie ankierów do badania ESS-CATI odbywało się w ramach czterech etapów. Pierwsze szkolenie przeprowadzono 3 tygodnie przed rozpoczęciem badania i poświęcono prezentacji projektu ESS (w tym „sposobom” zwiększania odsetka realizacji i zasadom konwersji miękkich odmów – *Refusal Aversion Training*, RAT) [Czaja, Blair 2005; Groves, Couper 1998; Groves, McGonagle 2001: 249–265]. Drugie szkolenie przeprowadzono 2 tygodnie przed rozpoczęciem realizacji badania. Zostały omówione poszczególne wersje kwestionariusza (skryptów), a następnie „pytanie po pytaniu” przeanalizowano skrypt w wersji A (jako „bazowy”). Trzecie szkolenie odbyło się w przeddzień rozpoczęcia terenowej realizacji projektu. Jego przedmiotem była powtórna, szczegółowa analiza kodów opisu rezultatu kontaktu, połączona z dyskusją dotyczącą ich „użyteczności” (ankierzy przeprowadzili ponad 140 wywiadów pilotażowych i mogli przedstawić własne doświadczenia). Ostatnie, czwarte z kolei szkolenie odbyło się w połowie okresu realizacji badania (4 tygodnie po rozpoczęciu). Miało ono na celu wymianę doświadczeń z realizacji badania oraz, co było ważniejsze, pełniło funkcje motywacyjne wobec ankierów.

4. Realizacja badania ESS-CATI trwała niemal 2 miesiące (od 14.12.2006 do 10.02.2007). Do jego realizacji wyznaczono 27. najbardziej doświadczonych ankierów, tzn. realizujących wywiady od kilku lat. Niezależnie od ich doświadczeń poddano ich wspomnianemu 4-etapowemu szkoleniu. W sumie podjęto ponad 14 000 kontaktów telefonicznych z ponad liczbą abonentów przekraczającą 1 400. Alokacja wywiadów dla ankierów w poszczególnych wersjach była losowa przy założeniu, iż jeden ankier nie może zrealizować więcej niż 35 wywiadów.

5. Nadzór i monitoring nad realizacją badania przebiegał dwutorowo. Był to wewnętrzny, bieżący nadzór sprawowany przez Instytut MillwardBrown SMG/KRC (odsłuchiwanie rozmów, obserwacja poprawności przebiegu sesji, zarówno od strony technicznej, jak i merytorycznej). Szczegółowej analizie poddano ponad 20% wywiadów zrealizowanych przez 23 ankierów (spośród 27). Kontrola każdego z przypadków obejmowała przynajmniej 75% całkowitego czasu trwania wywiadu. Instytut MillwardBrown SMG/KRC w odstępach tygodniowych miał obowiązek raportowania do IFiS PAN statystyk realizacyjnych. Drugi typ – nadzór zewnętrzny, sprawowany przez IFiS PAN – polegał na bieżącej, wrywkowej kontroli pracy ankierów (odsłuchiwanie rozmów, obserwacja poprawności przebiegu sesji itp.).

6. Niezależna kontrola terenowa (telefoniczna) przeprowadzona została przez IFiS PAN w lutym 2007 r. i objęła 10% przypadków odmów. Do kontroli

dobrano przypadki realizowane przez 3 ankieterów, którzy uzyskali najniższe wskaźniki realizacji, tzn. mieli relatywnie największą liczbę odmów.

Dość szczegółowo omówiono realizację terenową projektu JRA1. w Polsce. Wydaje się to konieczne, ponieważ analizujemy tutaj m.in. możliwości ograniczenia wpływu różnych sposobów kontaktowania się z respondentem na wynik badania.

### **EFEKT TECHNIKI (NA PRZYKŁADZIE ESS-PAPI I ESS-CATI)**

Jak wspomnieliśmy, różnice między techniką PAPI i CATI sprowadzają się do odmienności związanej z formułowaniem pytań, sposobem kontaktowania się z respondentem oraz sposobem komunikowania się z nim. Wszystko to skutkuje odmiennością efektów komunikacyjnych, a więc wpływa na uzyskiwane wyniki. Aby dowiedzieć się, jakie są różnice oraz czy istnieją możliwości redukcji efektu techniki i jakie one są, należy kolejno przeanalizować wszystkie trzy rodzaje odmienności.

#### **Różnice w sposobie sformułowania pytań**

Mówiąc o redukowaniu różnic dotyczących sposobu formułowania pytań, mamy na myśli podjęcie próby stworzenia narzędzia, a więc pytań kwestionariuszowych oraz instrukcji dla ankietera „odpornych” na efekt techniki. Największym „wyzwaniem” jest oczywiście przygotowanie kwestionariusza, różniącego się w zależności od technik, które zamierzamy wykorzystać w *Mixed Mode*, ale także ze względu na model *Mixed Mode* (sekwencyjny lub równoczesny). W przypadku modelu sekwencyjnego, po określeniu metody podstawowej („wyjściowej”) należy przygotować narzędzie dla tej techniki, a w kolejnym kroku stworzyć jego ekwiwalentne wersje dla innych technik. W drugim przypadku (równoczesny model *Mixed Mode*) brak jednej, „preferowanej” (podstawowej) metody. Wszystkie techniki traktowane są jako ekwiwalentne. Czym zatem się kierować i w jakiej kolejności przygotować taki kwestionariusz?. Możliwe są trzy podejścia [de Leeuw, Hox, Dillman 2008b: 312–313]:

1) pierwsze z nich zakłada optymalizację narzędzia dla danej techniki (tzw. *Mode Specific Design*) – oznacza to, że dla każdej techniki należy przygotować „osobne”, równoważne (niekoniecznie identyczne) narzędzie;

2) drugie umożliwia maksymalne upodobnienie każdego narzędzia na poziomie bodźca, tj. sformułowania pytania (tzw. *Uni-Mode Design*) [Dillman 2000: 232–240] oraz unifikację formatu pytań, odpowiedzi oraz instrukcji. Unifikacja

ta polega m.in. na: (a) zachowaniu jednolitej struktury pytań dla poszczególnych technik (zmiana struktury może zmienić bodziec); (b) redukowaniu liczby możliwych odpowiedzi (winny być takie same i włączone w treść pytania, także odpowiedź „Trudno powiedzieć”); (c) ustaleniu ekwiwalentnych instrukcji i reguł przejścia (jeśli byłyby bardzo skomplikowane, należy rozważyć restrukturyzację kwestionariusza w celu ich uproszczenia);

3) trzecie stawia na ekwiwalentność pytań na poziomie kognitywnym (tzw. *Generalized Mode Design*) [de Leeuw, Hox, Dillman, 2008b: 312–313]. Odchodzi się tutaj od literalnej unifikacji bodźca na poziomie formalnym, ponieważ identyczne sformułowanie nie zawsze zapewnia jego równoważność. Zakłada się natomiast, iż różne bodźce (pytania) mogą przynosić ekwiwalentne odpowiedzi na poziomie zapotrzebowania badawczego. Jeśli chcemy się dowiedzieć, czy skontaktowaliśmy się z właściwą osobą, to stosując technikę CATI zadamy pytanie: „Czy rozmawiam z.....? (imię i nazwisko)”; z kolei stosując ankietę Web-ową, postawimy pytanie: „Czy nazywasz się.....? (imię i nazwisko)” [de Leeuw 2009].

Wydaje się jednak słuszne, aby przy wyborze technik, które mają być łączone, kierować się jeszcze jedną zasadą. Gdy któraś z nich stawia przed nami większe w porównaniu z innymi „wymagania” w stosunku do kwestionariusza, sformułowania pytań, grafiki bądź języka programowania, to należy je uwzględnić w przypadku pozostałych technik. Jeśli w badaniach internetowych stosujemy technikę e-mailingu oraz ankietę Web-ową (bardziej „zaawansowaną”), to ta pierwsza, na poziomie formalnym, nie powinna być bardziej skomplikowana niż ankietę Web-owa – winny być przygotowane w podobny sposób, w podobnych „językach”.

W omawianym projekcie, podczas konstrukcji kwestionariusza ESS-CATI przyjęto perspektywę kognitywną, tj. zasadę unifikacji na poziomie bodźca, a więc ekwiwalentności koncepcyjnej, nie zaś unifikacji na poziomie formalnym. Oznacza to, iż dla „pierwotnego” kwestionariusza PAPI został stworzony jego „odpowiednik” dla techniki CATI. Poniżej prezentujemy przykład pytania źródłowego (wersja PAPI) i pytania adaptowanego do CATI (takiego, jakie widział ankietę na ekranie komputera).

Przykład pytania: ESS-PAPI

|  |    |    |    |    |    |    |    |    |    |                          |    |                     |
|--|----|----|----|----|----|----|----|----|----|--------------------------|----|---------------------|
| <p>B24. Biorąc wszystko pod uwagę, ogólnie, na ile jest P. zadowolony/-a ze swojego obecnego życia? Odpowiadając, proszę <u>posłużyć się kartą [KARTA 10]</u>, gdzie 0 oznacza, że jest P. zupełnie niezadowolony/-a, a 10 oznacza całkowicie zadowolony/-a.</p> |    |    |    |    |    |    |    |    |    |                          |    |                     |
| <p>KARTA 10. POZOSTAWIĆ DO B27</p>   |    |    |    |    |    |    |    |    |    |                          |    |                     |
| Zupełnie niezadowolony/-a  |    |    |    |    |    |    |    |    |    | Całkowicie zadowolony/-a |    | (Trudno powiedzieć) |
| 00   | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10                       | 88 |                     |

Przykład pytania: ESS-CATI

|   |    |    |    |    |    |    |    |    |    |                         |    |                     |
|---|----|----|----|----|----|----|----|----|----|-------------------------|----|---------------------|
| <p>TB24. Biorąc wszystko pod uwagę, ogólnie, na ile jest P. zadowolony/-a ze swojego obecnego życia? Odpowiadając, proszę <u>wskazać liczbę od 0 do 10</u>, gdzie 0 oznacza, że jest P. zupełnie niezadowolony/-a, a 10 oznacza, że całkowicie zadowolony/-a. <u>Pozostałe liczby służą do wyrażenia opinii pośrednich.</u></p> |    |    |    |    |    |    |    |    |    |                         |    |                     |
| Zupełnie niezadowolony/-a   |    |    |    |    |    |    |    |    |    | Całkowicie zadowolony/a |    | (Trudno powiedzieć) |
| 00  | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10                      | 88 |                     |

Powyższy przykład stanowi dobrą ilustrację odmienności sformułowania pytania w PAPI i CATI oraz instrukcji do pytania, w tym wykorzystania materiałów pomocniczych (chodzi o Kartę Respondenta).

### Różnice w sposobie kontaktowania się i ich efekty

Różnice w sposobie kontaktowania się najlepiej obrazuje odsetek realizacji próby wylosowanej przy zastosowaniu wspomnianych dwóch technik: wywiadu osobistego (PAPI/CAPI) oraz telefonicznego (CATI). Dla porównania, dla poszczególnych krajów biorących udział w projekcie JRA 1, w poniższych tabelach 1. i 2.) zamieszczono także odsetek realizacji próby wylosowanej (*Response Rate: RR*) w badaniu zasadniczym<sup>12</sup>.

<sup>12</sup> Odsetek ten został wyliczony zgodnie z definicją tzw. Response Rate (RR2), przyjętą przez American Association for Public Opinion Research [2008]. Jest on definiowany jako liczba zrealizowanych wywiadów (kompletnych lub częściowo zrealizowanych), dzielona przez liczebność wylosowanej próby, pomniejszoną o liczbę przypadków, które nie powinny znaleźć się w próbie, np. błędny adres/numer telefonu; adres/numer telefonu instytucji itp.

TABELA 1. Odsetek realizacji próby (AAPOR RR2) w poszczególnych krajach biorących udział w projekcie JRA1 dla badania realizowanego techniką PAPI/CAPI oraz CATI

| Kraj       | Próba        | Technika ESS r. 3: (RR%) | ESS r. 3 CATI: (RR%) | Różnica |
|------------|--------------|--------------------------|----------------------|---------|
| Cypr       | Gospodarstw  | PAPI: 67,3               | 8,3                  | -59,0   |
| Niemcy     | Indywidualna | CAPI: 52,9               | 24,3                 | -28,6   |
| Węgry      | Adresowa     | PAPI: 65,9               | 25,2                 | -40,7   |
| Polska     | Indywidualna | PAPI: 70,1               | 35,3                 | -34,8   |
| Szwajcaria | Gospodarstw  | CAPI: 50,0               | 41,5                 | -8,6    |

Źródło: Roberts et al.: 2010.

Dla poszczególnych krajów ogólny odsetek realizacji próby wylosowanej jest, co oczywiste, różny – w zależności od wersji kwestionariusza ESS-CATI (A, B, C), wyróżnionej ze względu na przewidywany czas trwania wywiadu. Ilustruje to tabela 2.

TABELA 2. Odsetek realizacji próby wylosowanej w zależności od wersji kwestionariusza ESS-CATI

| Kraj       | Technika ESS r. 3 (RR%) | ESS r. 3 CATI (RR%) | Wersja A (%) | Wersja B (%) | Wersja C (%) |
|------------|-------------------------|---------------------|--------------|--------------|--------------|
| Cypr       | PAPI: 67.3              | 8,3                 | 8,8          | 6,2          | 9,0          |
| Niemcy     | CAPI: 52.9              | 24,3                | 21,8         | 26,3         | 25,7         |
| Węgry      | PAPI: 65.9              | 25,2                | 19,3         | 24,5         | 31,5         |
| Polska     | PAPI: 70.1              | 35,3                | 34,1         | 38,0         | 35,3         |
| Szwajcaria | CAPI: 50.0              | 41,5                | 37,9         | 39,5         | 50,2         |

Źródło: Roberts et al. 2010.

Jak można zauważyć, w każdym z krajów uczestniczących w badaniu ESS-CATI uzyskano znacząco niższe odsetki realizacji niż w badaniu zasadniczym (PAPI lub CAPI). Największe różnice – blisko 60 punktów procentowych – odnotowano dla Cypru (PAPI vs CATI), najmniejsze dla Szwajcarii, niespełna 9 punktów procentowych (CAPI vs CATI). Co ciekawe, uzyskano tam – w porównaniu z innymi krajami – najniższy odsetek realizacji w badaniu zasadniczym, a najwyższy w badaniu przy zastosowaniu techniki CATI. Zwraca także uwagę większa rozpiętość odsetka realizacji w badaniu CATI niż w badaniu zasadniczym: o ile dla technik *face-to-face* wynosi ona 20 punktów procentowych, o tyle dla CATI osiąga ponad 33 punkty procentowe. Nie podejmujemy się interpretacji tak



dużych rozpiętości wyników, ponieważ wymagają one znajomości kulturowych i społecznych uwarunkowań realizacji badań w każdym z krajów<sup>13</sup>.

W Polsce zrealizowano 353 wywiady, przy czym w 301 przypadkach w całości, a w 52 częściowo (zostały przerwane przez respondenta lub respondenci nie wyrazili zgody na realizację drugiej części wywiadu: wersja C). Spośród wszystkich wywiadów w wersji A udało się przeprowadzić 136, w wersji B – 76, a w wersji C – 141. Pamiętając, iż proporcje w próbie wynosiły odpowiednio: 400, 200 i 400, zauważyć można, że najwięcej wywiadów udało się zrealizować w wersji B (najkrótszej – założony czas trwania 45 min.), następnie w wersji C (2×30 min.), natomiast najmniej w wersji pełnej A (60 min.). To, że w wersji C zrealizowano relatywnie mało wywiadów (tylko o 5 więcej, niż w wersji A) wydawać się może zaskakujące, gdyż w założeniu miała to być najbardziej „przyjazna” dla respondentów opcja. Jednocześnie respondenci aż w 41 przypadkach po zakończeniu realizacji pierwszej części wywiadu nie wyrazili zgody na jego kontynuację. Wyjaśnić to można, jak się wydaje, błędną aranżacją do tej wersji kwestionariusza. W projekcie przyjęto, że w trakcie aranżacji wstępnej do pierwszej części wywiadu (część a.) respondent jest informowany, że wywiad będzie trwał ok. 30 min., a dopiero po jego zakończeniu, w aranżacji końcowej, proszony jest o zgodę na drugi wywiad, także trwający 30 min. (część b.). W takiej sytuacji część respondentów czuła się „oszukana” i nie wyrażała zgody na realizację drugiej części wywiadu.

Warto wspomnieć, że rozpiętość czasów trwania wywiadów była dość duża. Dla poszczególnych wersji maksymalna i minimalna długość wywiadu (podczas jednego kontaktu) wynosiła odpowiednio: wersja A – 150 min. i 41 min., wersja B – 75 min. i 32 min., wersja C (część a. i b.) – 132 min. i 44 min.

Jeśli chodzi o wywiady niezrealizowane, to bardzo liczną kategorię stanowiły przypadki „braku odpowiedzi po 7 sygnałach”, a więc te, w których nikt nie odbierał telefonu (120 na ogółem 131 przypadków „nienawiązania kontaktu”). Wydaje się, że było to związane z terminem realizacji badania (14.12.2006–10.02.2007). Był to okres wyjazdów „świętecznych” i „międzyświętecznych”, a potem „ferii zimowych” w szkołach.

Z kolei w sytuacji, gdy ankieterowi udało się nawiązać kontakt, najczęstsze były przypadki, w których nie zakończył procedury doboru respondenta i umawiał się na powtórny telefon w innym terminie (25 przypadków na 59). Sytuację tę traktować można jako „pośrednią odmowę”, podobnie 13 przypadków, w któ-

<sup>13</sup> Zgodnie z interpretacją Dominique Joye z Université de Lausanne, narodowego koordynatora projektu ESS, w Szwajcarii „większość spraw załatwia się przez telefon”. Stąd relatywnie niski odsetek realizacji wywiadu techniką CAPI, a wysoki – CATI.

rych wylosowano respondenta, ale nie można było się z nim skontaktować lub wielokrotnie „przekładał” terminy wywiadu.

Bardzo liczną kategorię stanowiły „odmowy” udzielenia wywiadu (45,7% przypadków, w których nawiązano kontakt z numerem należącym do próby). Należy jednak zwrócić uwagę, że na 368 takich przypadków jedynie w 63. zgody na wywiad nie wyraził wylosowany respondent (w 305 przypadkach „odmowa” miała miejsce przed wylosowaniem respondenta). Oznacza to, że kluczową sprawą dla powodzenia badania jest pozyskanie kooperacji osoby, która odbiera telefon.

Z kolei zasadniczą przyczyną konieczności włączenia dodatkowych numerów do pierwotnie wylosowanej próby był „nieistniejący numer” (ponad 70% wszystkich dodatkowych numerów) oraz „numery instytucji” (24,3%).

Mówiąc o różnicach w sposobie kontaktowania się (kontakt telefoniczny – kontakt *face-to-face*), kolejną ważną kwestią jest pytanie o podobieństwo struktury próby zrealizowanej w badaniu ESS-PAPI i ESS-CATI. Podstawowe charakterystyki społeczno-demograficzne dla obu prób są zawarte w tabeli 3.

TABELA 3. Porównanie próby zrealizowanej w ESS-PAPI i ESS-CATI

| Polska: ESS r. 3                         | PAPI (%)    | CATI (%)   | Różnica |
|--|-------------|------------|---------|
| N  | 1718 (100%) | 353 (100%) |         |
| Mężczyźni [w %]                          | 47,4        | 40,1       | 7,3 *   |
| Średnia wieku [w latach]                 | 43,7        | 47,7       | 4,0     |
| Obecnie wykonuje pracę odpłatną [w %]    | 48,0        | 51,0       | 3,0     |
| Miejsce zamieszkania                     |             |            |         |
| - duże miasto                            | 24,9        | 16,6       | 8,3     |
| - przedmieścia lub obrzeża dużego miasta | 5,3         | 9,6        | 4,3     |
| - średnie lub małe miasto                | 31,3        | 32,1       | 0,8     |
| - wieś                                   | 37,6        | 33,4       | 4,2     |
| - pojedyncze gospodarstwo lub dom na wsi | 0,9         | 8,3        | 7,4     |
| Główne zajęcie w ciągu ostatnich 7 dni   |             |            |         |
| - praca odpłatna                         | 35,8        | 45,2       | 9,4 *   |
| - nauka                                  | 14,5        | 6,6        | 7,9     |
| - bezrobotny, poszukujący pracy          | 3,9         | 0,6        | 3,3     |
| - bezrobotny, nie poszukujący pracy      | 2,8         | 1,4        | 1,4     |
| - trwała choroba, niepełnosprawność      | 1,6         | 0,3        | 1,3     |
| - emerytura/renta                        | 25,3        | 32,6       | 7,3     |
| - opieka nad domem                       | 14,9        | 10,1       | 4,8     |
| - inne                                   | 1,1         | 3,2        | 2,1     |
| Lata skolaryzacji (średnia)              | 12,4        | 13,0       | 0,6     |

\*  $p \leq 0,05$

Źródło: Roberts et al. 2010 i obliczenia własne

Podsumowując, różnice w efektach kontaktowania się *face-to-face* oraz telefonicznego są znaczące. Czy zatem, w badaniu ESS w Polsce technika CATI może stanowić alternatywę dla PAPI? Tak, jeśli uznamy za satysfakcjonujące, że: (i) ogólny odsetek zrealizowanych wywiadów nie jest zbyt niski (35%); (ii) w warunkach polskich koszty uzyskania 1 przypadku w badaniu CATI są niewiele niższe niż PAPI i wyrażają się stosunkiem 9: 10. Należy jednak pamiętać, że odpowiedź ta jest warunkowa i stanowi odpowiedź częściową. Nie wiemy, jaki jest wpływ zastosowanej techniki (chodzi o tzw. *mode effects*) na wartość uzyskanych danych, a więc efekty komunikacyjne. Czy i w jakim stopniu uzyskane wyniki w ESS-CATI różnią się od wyników ESS-PAPI spróbujemy odpowiedzieć w dalszej części opracowania.

### Różnice w sposobie komunikowania się i ich efekty

Obecnie na przykładzie kilku wybranych pytań, różnych ze względu na przedmiot (postawy/opinie i zachowania) oraz strukturę (dychotomiczne oraz skale semantyczne i liczbowe) przeanalizujemy, czy istnieją różnice efektów komunikacyjnych, a więc dotyczące udzielanych odpowiedzi w wywiadzie ESS-PAPI i ESS-CATI, a jeśli występują, to czy są znaczące?

Porównanie wyników tych dwóch badań prowadzić będziemy na próbach „matchowanych”, tzn. porównywalnych ze względu na kryteria warstwowania (płeć, wiek respondenta oraz wielkość miejscowości zamieszkania). W celu zrównoważenia struktur obu prób przeważaliśmy próbę zrealizowaną w badaniu ESS-PAPI (znacząco większą) do struktury próby zrealizowanej w badaniu ESS-CATI. Dalsze próby zrównoważenia struktur porównywanych prób ograniczała stosunkowo mała liczebność próby zrealizowanej w badaniu ESS-CATI.

Jako pierwsze przeanalizujemy pytanie dychotomiczne, o zachowania, związane z udziałem w wyborach do sejmiku we wrześniu 2005 r. W tabeli 4 przedstawiono rozkład odpowiedzi.

TABELA 4. Rozkład odpowiedzi na pytanie o udział w wyborach do sejmiku w 2005 r.

| TB11/B11. Obecnie wielu ludzi z różnych przyczyn nie bierze udziału w wyborach. Czy brał/-a P. udział w ostatnich wyborach do sejmiku we wrześniu 2005 r.? | PAPI (%) | CATI (%) | Różnica |
|--|----------|----------|---------|
| 1. Tak   | 64,3     | 74,7     | 10,4    |
| 2. Nie   | 30,8     | 24,0     | 6,8     |
| 3. Nie byłem/-am uprawniony/-a do głosowania   | 4,4      | 1,3      | 3,1     |
| 8. trudno powiedzieć   | 0,4      | 0,0      | 0,4     |
| 9. brak odpowiedzi   | 0,1      | 0,0      | 0,1     |

$$\text{Chi}^2 = 7,58; p \leq 0,01; \text{df} = 1$$

Różnice rozkładów odpowiedzi (z wyłączeniem odpowiedzi: „nie byłem uprawniony”/„trudno powiedzieć”)<sup>14</sup> są istotne w teście Chi<sup>2</sup>. Należy zwrócić uwagę na ponad 10-punktową różnicę w przypadku odpowiedzi potwierdzającej udział w wyborach: w wywiadzie CATI respondenci częściej udzielali takich odpowiedzi niż w PAPI.

Warto przypomnieć, że frekwencja wyborcza w 2005 r. wyniosła 40,6%, a więc także w wywiadzie PAPI odsetek odpowiedzi twierdzących jest zawyżony w porównaniu ze stanem faktycznym. Można wobec tego przypuszczać, że jest to pytanie drażliwe, skłaniające do udzielania społecznie akceptowanych odpowiedzi, choć – co zaskakujące – efekt ten ujawnił się mocniej w CATI. Być może taki wynik jest powiązany ze wspomnianą wcześniej tendencją do „potakiwania”, a może jest to efekt różnic doboru próby i domniemanego jej „wypaczenia” w ESS-CATI.

W następnej kolejności przeanalizujemy pytanie dychotomiczne o postawy dotyczące poczucia związku z partią polityczną. W tabeli 5. ujęto rozkład odpowiedzi na to pytanie.

TABELA 5. Rozkład odpowiedzi na pytanie o poczucie związku z jakąś partią polityczną

| TB20a/B20a. Czy istnieje jakaś partia polityczna, z którą czuje się P. bardziej związany/-a niż z innymi partiami, jest P. jej sympatykiem? | PAPI (%) | CATI (%) | Różnica |
|---|----------|----------|---------|
| 1. Tak  | 26,7     | 47,0     | 20,3    |
| 2. Nie  | 71,9     | 53,0     | 18,9    |
| 8. trudno powiedzieć  | 1,3      | 0,0      | 1,3     |
| 9. brak odpowiedzi  | 0,2      | 0,0      | 0,2     |

$$\text{Chi}^2 = 47,96; p \leq 0,05; \text{df} = 1$$

Wynik testu Chi<sup>2</sup> także wskazuje na istotne różnice w rozkładach odpowiedzi na to pytanie. Analogicznie jak w poprzednim pytaniu, w wywiadzie CATI występuje znacząca przewaga odpowiedzi twierdzących (w porównaniu z wywiadem PAPI). Mimo iż pytania te różni przedmiot – jedno z nich dotyczy zachowań, drugie opinii – w obu prawdopodobnie ujawnia się efekt, wspomnianej wcześniej, tendencji do „potakiwania”.

<sup>14</sup> Za wyłączeniem tych kategorii przemawia to, że zwiększanie liczby kategorii obniża wiarygodność testu dla rozkładów (tym bardziej, że liczebności dla tych kategorii są bardzo niskie). Powyższa uwaga dotyczy również kolejnych tabel.

Pytanie trzecie o opinie (postawy), dotyczy zaufania do innych ludzi. Wykorzystano w nim 11-punktową skalę liczbową z opisanymi krańcami. W tabeli 6. zilustrowano rozkład odpowiedzi na to pytanie.

TABELA 6. Rozkład odpowiedzi na pytanie o poczucie zaufania

| TA8/A8. Czy, ogólnie biorąc, uważa P., że większości ludzi można ufać, czy też, że w kontaktach z ludźmi ostrożności nigdy za wiele? Odpowiadając proszę wskazać liczbę od 0 do 10, gdzie 0 oznacza, że ostrożności nigdy za wiele, a 10 oznacza, że większości ludzi można ufać. Pozostałe liczby służą do wyrażenia opinii pośrednich. | PAPI (%) | CATI (%) | Różnica |
|--|----------|----------|---------|
| 00. Ostrożności nigdy za wiele   | 10,3     | 20,0     | 9,7     |
| 01.  | 6,6      | 3,0      | 3,6     |
| 02.  | 11,6     | 7,7      | 3,9     |
| 03.  | 14,7     | 7,3      | 7,4     |
| 04.  | 8,8      | 10,3     | 1,5     |
| 05.  | 24,9     | 26,0     | 1,1     |
| 06.  | 6,7      | 7,3      | 0,7     |
| 07.  | 7,2      | 7,0      | 0,2     |
| 08.  | 5,8      | 8,3      | 2,5     |
| 09.  | 1,1      | 1,7      | 0,6     |
| 10. Większości ludzi można ufać  | 1,8      | 1,3      | 0,5     |
| 88 trudno powiedzieć   | 0,5      | 0,0      | 0,5     |

$\chi^2 = 44,34; p \leq 0,01; df = 10$

Różnice rozkładów odpowiedzi dla tego pytania także są istotne w teście  $\chi^2$ . Taki wynik można wiązać z trudnością odbioru skali liczbowej w wywiadzie CATI. W wywiadzie PAPI, po odczytaniu treści pytania, ankieter wręczał respondentowi kartę, na której zamieszczono wizualizację skali, tzn. 11 punktów, oznaczonych liczbami od „00” do „10”, z odpowiednim opisem jej krańców: „Ostrożności nigdy za wiele” oraz „Większości ludzi można ufać”. Respondent widział skalę, na której ma wskazać odpowiedni punkt oraz opis jej krańców. W wywiadzie CATI natomiast respondent musiał „wyobrazić sobie” taką skalę jedynie na podstawie odczytanego przez ankietera opisu. Bez wątpienia musiało być to dla niego trudne zadanie. Analizując rozkłady odpowiedzi w wywiadzie CATI, zaobserwować można, że respondenci „skracali” skalę i wskazywali odczytywany jako pierwszy punkt skrajny („00. Ostrożności nigdy za wiele”)

oraz środkowy („05”) [Sztabiński F. 2003: 147–175]. W sumie te dwa punkty wskazało 46% respondentów<sup>15</sup>. O występowaniu tendencji do „spłaszczania” skali w wywiadzie CATI świadczy także niska częstość wskazań punktów „pośrednich”, w szczególności punktów: „01”, „02” i „03”. Jeden z tych punktów w wywiadzie PAPI wskazała prawie 1/3 respondentów (32,9%), zaś w wywiadzie CATI tylko 18%<sup>16</sup>.

Ostatnie pytanie o postawy wobec osób mających orientację homoseksualną, z wykorzystaniem 5-punktowej skali semantycznej. W tabeli 7 zilustrowano rozkład odpowiedzi na to pytanie.

TABELA 7. Rozkład odpowiedzi na pytanie o stosunek do osób o orientacji homoseksualnej

| TB31/B31. Proszę powiedzieć, w jakim stopniu zgadza się P. albo nie zgadza z następującym stwierdzeniem: „Osoby o orientacji homoseksualnej, geje i lesbijki, powinni mieć swobodę układania sobie życia wg własnych przekonań”. Czy P. ... <i>ODCZYTAĆ</i> | PAPI (%) | CATI (%) | Różnica |
|---|----------|----------|---------|
| 1. Zdecydowanie się zgadzam   | 8,9      | 6,7      | 2,2     |
| 2. Zgadzam się  | 32,6     | 37,0     | 4,4     |
| 3. Ani się zgadzam, ani nie zgadzam   | 20,1     | 23,0     | 2,9     |
| 4. Nie zgadzam się  | 20,9     | 19,7     | 1,2     |
| 5. Zdecydowanie się nie zgadzam   | 10,3     | 12,3     | 2,0     |
| 7. odmowa odpowiedzi  | 0,1      | 0,0      | 0,1     |
| 8. trudno powiedzieć  | 7,1      | 0,0      | 7,1     |
| 9. brak odpowiedzi  | 0,1      | 1,3      | 1,2     |

$$\text{Chi}^2 = 4,04, p < 0,05 (0,4008122); df = 4$$

Generalnie różnice rozkładów odpowiedzi nie są duże (maksymalna wynosi |4,4| punktu procentowego). Wynik testu  $\text{Chi}^2$  – przy założeniu, że  $p < 0,05$  – nie wykazał istotnych różnic w rozkładach odpowiedzi. Jest to zupełnie inna sytuacja niż w pytaniu poprzednim, mimo że obydwa dotyczyły opinii. Wydaje się,

<sup>15</sup> Ankieterzy realizujący badanie zwracali uwagę na to, iż skale te sprawiały respondentom wiele trudności: prosili oni o powtarzanie opisów ich krańców lub dodatkowe wyjaśnienia. Zaobserwowali także tendencję do ich „skracania” i „przekładania” na skale słowną (tzn. łączenia poszczególnych punktów w większe całości i wskazywania określonych punktów, odpowiadających w ich odczuciu skali słownej).

<sup>16</sup> Tendencji takiej nie zaobserwowano w przypadku „drugiej połówki” skali (punkty: od „06” do „10. Większości ludzi można ufać”). Jest to prawdopodobnie powiązane ze skośnością rozkładu odpowiedzi – punkty te wskazywane były relatywnie rzadziej – w sumie wskazało je 22% respondentów w wywiadzie PAPI i 25% w CATI.

iż można to wyjaśnić dwojako. Po pierwsze, pięciopunktowa skala semantyczna była dla respondentów z pewnością „łatwiejsza” niż 11-punktowa skala liczbowa. Po drugie, powyższe pytanie dotyczyło drażliwej i bardzo konkretnej kwestii: akceptacji bądź jej braku dla związków homoseksualnych, natomiast poprzednie sprawy bardzo ogólnej, związanej z indywidualnymi doświadczeniami – zaufania do ludzi. W wywiadzie PAPI respondenci mieli możliwość zastanowienia się, „głośnego” myślenia i dopiero potem wskazania właściwego punktu na skali. W wywiadzie CATI, ze względu na charakter i szybkie tempo rozmowy telefonicznej, byli oni pozbawieni takiej możliwości.

Analizując efekt odmienności sposobów komunikowania się w PAPI i CATI, z pewnością można stwierdzić, iż wystąpił on w analizowanych pytaniach. Zaskoczenie stanowi to, że szczególnie silnie wystąpił on w przypadku pytań dychotomicznych (różnice sięgają 20 punktów procentowych), przy czym, co oczywiste, jest on słabszy w pytaniach dotyczących zachowań, silniejszy natomiast w pytaniach o opinie. W przypadku pytań dychotomicznych różnice w odpowiedziach wynikają najprawdopodobniej ze wspomnianej tendencji do „potakiwania” w sytuacji wywiadu CATI.

W przypadku pytań ze skalami nie można sformułować tak jednoznacznego wniosku. Przede wszystkim zależy to od rodzaju zastosowanej skali. Stosując abstrakcyjne, długie i z pewnością trudne dla respondentów skale (np. liczbowe) musimy liczyć się z możliwością wystąpienia efektu techniki. W przypadku prostych skal semantycznych ma on raczej ograniczone znaczenie. Jak wskazaliśmy, czynnikiem „interweniującym” może być także tematyka pytania.

Mówiąc o efekcie techniki, warto zwrócić uwagę na jeszcze jedną kwestię. Otóż w żadnym z analizowanych pytań w wywiadzie CATI nie pojawiły się odpowiedzi typu: „nie wiem”/”trudno powiedzieć”, spotykane w wywiadach PAPI. Ich odsetek nie jest wysoki, oscyluje wokół 0,5–1%. Wyjątek stanowi zadane w wywiadzie PAPI pytanie dotyczące związków homoseksualnych, w którym odsetek tego typu odpowiedzi wyniósł ponad 7% (przypomnijmy, w CATI: 0,0%, a różnica ta jest istotna statystycznie). Interpretując ten wynik, znowu należy odwołać się do specyfiki rozmowy telefonicznej, w której odpowiedź „nie wiem” może zostać odebrana jako wyraz zniechęcenia czy znudzenia. W bezpośredniej rozmowie taka odpowiedź po „chwili namysłu”, z ewentualnym uzasadnieniem, nie zostanie zinterpretowana w ten sposób.

### **Wpływ różnic związanych z odmiennością sposobu komunikowania się na związek między zmiennymi**

Różnice rozkładów odpowiedzi na poszczególne pytania nie przesądzają o sile związków i ewentualnych zależnościach z innymi zmiennymi. Stąd też należy postawić pytanie o wpływ różnic techniki zbierania danych (odmienności sposobu komunikowania się) na związki z innymi zmiennymi, a więc wynik. Gdyby okazało się, że korelacje między odpowiedziami respondentów na pytanie TX a ich odpowiedziami na pytanie TY w wywiadzie CATI oraz analogicznymi uzyskanymi w wywiadzie PAPI (tzn. X i Y) różnią się między sobą w istotny sposób, to wówczas można zakwestionować trafność wniosków odwołujących się do zależności między tymi zmiennymi.

Dla pytań dychotomicznych wyznaczyliśmy równania regresji logistycznej [deMaris, 1992] i porównaliśmy wartości współczynnika „b”. Dla pytania o udział w wyborach parlamentarnych wprowadziliśmy zestaw 4 zmiennych niezależnych: płeć, wiek, miejsce zamieszkania oraz wykształcenie. Ze względu na niskie liczebności zmiennych niezależnych wprowadzanych do modelu dla danych z badania ESS-CATI, zrekodowaliśmy je do zmiennych dwuwartościowych w następujący sposób: wiek (15–44; 45+ lat), miejsce zamieszkania (miasto; wieś) i wykształcenie (dwie zmienne: (i) średnie i poniżej średniego vs pozostali; (ii) powyżej średniego vs pozostali). W analogiczny sposób postąpiliśmy w odniesieniu do pytania o poczucie związku z partią polityczną, tzn. wprowadziliśmy te same zestawy zmiennych niezależnych: płeć, wiek i miejsce zamieszkania.

Dla pytań o poczucie zaufania i stosunek do osób o orientacji homoseksualnej wyznaczyliśmy równania regresji liniowej i porównaliśmy wartości współczynnika „b”. Dla obu pytań wprowadziliśmy te same zestawy 4 zmiennych: płeć, wiek, miejsce zamieszkania oraz wykształcenie<sup>17</sup>. Są one zgodne pod względem konstrukcji, tzn. zmienne zostały odpowiednio zrekodowane, tak, by w obu zbiorach przyjmowały te same wartości. Wyniki analizy zamieszczono w tabeli 8.

---

<sup>17</sup> Mówiąc o wykształceniu, mamy na myśli poziom wykształcenia określony wg kategorii ISCED, ustalony na podstawie pytania F6/TF6 (9 kategorii).



TABELA 8. Wartości metrycznych współczynników regresji logistycznej dla pytań TB11/B11 i TB20a/B20a oraz regresji liniowej dla pytań TA8/A8 i TB3 /B31

| Pytanie  | Zmienna wyjaśniająca | ESS-PAPI |           | ESS-CATI |           |
|--|----------------------|----------|-----------|----------|-----------|
|  |                      | „b”      | istotność | „b”      | istotność |
| Pyt. TB11/B11. Udział w wyborach                             | płeć                 | -.019    | .865      | -.558    | .064      |
|  | wiek                 | .535     | .000      | .905     | .002      |
|  | miejsce zamieszkania | -.094    | .402      | .176     | .545      |
|  | wykształcenie        | .889     | .000      | .538     | .064      |
| Pyt. TB20a/B20a. Poczucie związku z partią                   | płeć                 | -.513    | .000      | -.365    | .131      |
|  | wiek                 | .257     | .031      | .647     | .010      |
|  | miejsce zamieszkania | -.366    | .002      | -.255    | .298      |
| Pyt. TA8/A8. Poczucie zaufania                               | płeć                 | -.039    | .743      | .617     | .055      |
|  | wiek                 | -.062    | .064      | -.440    | .174      |
|  | miejsce zamieszkania | -.011    | .920      | .053     | .601      |
|  | wykształcenie        | .293     | .000      | .158     | .105      |
| Pyt. TB31 /B31. Stosunek do osób o orientacji homoseksualnej | płeć                 | -.017    | .766      | -.135    | .327      |
|  | wiek                 | .120     | .000      | .532     | .000      |
|  | miejsce zamieszkania | .213     | .000      | .019     | .657      |
|  | wykształcenie        | -.138    | .000      | .018     | .667      |

Porównanie wartości niestandardyzowanych metrycznych współczynników  $b$  regresji logistycznej i liniowej wskazuje jedynie, że dla analizowanych pytań odpowiedzi respondentów są inaczej powiązane z podstawowymi zmiennymi społeczno-demograficznymi w badaniu dla ESS-PAPI, a inaczej w badaniu ESS-CATI. Niewątpliwie taki wynik jest konsekwencją dysproporcji wielkości prób w badaniu ESS-PAPI ( $N = 1721$ ) oraz ESS-CATI ( $N = 300$ ) i może mieć wpływ na to co jest istotne a co nie jest. Obie próby są zatem różne i to w takim stopniu, że inne zmienne pozwalają przewidywać wartości zmiennej zależnej. Innymi słowy, w oparciu o wyniki tak samo skonstruowanych równań regresji w każdym z badań otrzymujemy odmienne układy zależności i w konsekwencji uzyskujemy różne wyniki.

### ESS-PAPI/CATI MIXED MODE: ZA I PRZECIW

Możliwości i warunki prowadzenia badań surveyowych na próbach losowych ulegają coraz większym ograniczeniom, stają się coraz trudniejsze. Związane jest to przede wszystkim ze wzrostem wiedzy, a właściwie uproszczonych wyobrażeń o badaniach, wzrostem podmiotowości i poczucia prywatności respondentów, ale także wzrostem poczucia zagrożenia bezpieczeństwa osobistego. Nie bez

znaczenia jest także obiektywny wzrost „obciążeń i zobowiązań” zawodowych [Sztabiński F., Sztabiński P.B. 1994: 141–164, Sztabiński P.B. et al. 2008: 39–84]. Wszystko to skutkuje obniżającym się odsetkiem realizacji próby i błędem braku odpowiedzi (*unit nonresponse*). Obecnie, w standardowych badaniach konsumenckich i opinii publicznej, współczynnik realizacji (przy zastosowaniu PAPI/CAPI) kształtuje się w granicach 20–40%. W tej sytuacji *Mixed Mode Survey Design* (rozumiany jako procedura zbierania tych samych informacji różnymi technikami od różnych osób) może się stać swoistym remedium. Zakłada się tutaj, że wylosowane osoby mają pewne preferencje dotyczące formy kontaktu i sposobu komunikowania, pewne techniki są – o czym wspominaliśmy – bardziej lub mniej adekwatne ze względu na społeczne i psychologiczne charakterystyki badanych.

Jeśli przyjmiemy, że *Mixed Mode Survey Design* może redukować błąd braku odpowiedzi i błąd pokrycia, to musimy także przyjąć, iż może generować błąd pomiaru, związany z techniką (*Mode Effects*). Analizując wyniki badania ESS-PAPI oraz ESS-CATI r.3 dla Polski (ale także dla innych krajów), potwierdzamy, że tak jest.

Po pierwsze, odsetek realizacji wywiadu ESS-CATI (niezależnie od wersji) w niektórych krajach jest blisko o połowę niższy niż przy zastosowaniu PAPI (Polska: 35 vs 70%) czy CAPI (Niemcy: 53 vs 24%). W przypadku innych krajów jest on znacznie niższy: Węgry, Cypr (odpowiednio: 65,9 vs 25,2% oraz: 67,3 vs 8,3%). Wyjątek stanowi Szwajcaria, która w ESS-CAPI uzyskała odsetek realizacji 50%, a w ESS-CATI nieco niższy – 41,5%. Wynik ten zdaje się wskazywać na problem różnego poziomu akulturacji technik. Pierwszym podstawowym wnioskiem jest zatem to, iż CATI, biorąc pod uwagę sposób kontaktowania się z respondentem, nie stanowi „konkurencji” dla PAPI (przynajmniej w Polsce), ale także dla CAPI (np. Niemcy), co potwierdza występowanie efektu techniki, przynajmniej w sferze kontaktu.

Po drugie, analizując uzyskane rezultaty (rozkłady odpowiedzi) zauważyć można, iż różnią się one między sobą. W większości przypadków są to różnice istotne statystycznie i zależą nie tylko od struktury pytania, lecz także od jego przedmiotu. W tym przypadku nie sposób określić, które z uzyskanych wyników (CATI czy PAPI) są „prawdziwe”, nie o to zresztą chodzi. Następny podstawowy wniosek jest taki, że CATI przynosi odmienne wyniki (chodzi o odpowiedzi respondentów) w porównaniu do PAPI, przynajmniej w Polsce. Potwierdza to wystąpienie efektu techniki w sferze komunikacji.

Po trzecie, porównując wnioski, które można sformułować na podstawie badania PAPI i CATI (chodzi o związki między zmiennymi), znowu trzeba uznać,

iż są one odmienne. Jeśli bierzemy pod uwagę poszczególne cechy społeczno-demograficzne (płeć, wiek itd.), to ich związek z niektórymi zmiennymi zależnymi jest podobny. Nie zmienia to ogólnego wniosku dotyczącego wystąpienia efektu techniki.

Sformułowane wnioski mają oczywiście bardzo ograniczony charakter, głównie ze względu na specyfikę zebranego materiału, który jest pochodną celów i założeń przyjętych w projekcie JRA1. Ograniczenia dotyczą przede wszystkim wielkości próby założonej do badania ESS-CATI (próba N = 1000 realizowana do wyczerpania, w podziale na 3 podpróby, w wyniku czego podstawy procentowania były bardzo niskie).

Czy zatem zastosowanie *Mixed Mode Survey Design* redukuje błędy pokrycia, braku odpowiedzi i zwiększa błąd pomiaru, głównie za sprawą efektu techniki? Wydaje się, że procedura ta może redukować te błędy, ale . . . W chwili obecnej wszystkie pytania dotyczące *Mixed Mode Survey Design* wydają się pozostawać bez jednoznacznej odpowiedzi. Nie jest on gotowym „przepisem”, umożliwiającym redukcję błędów braku odpowiedzi. Musimy zawsze brać pod uwagę czasowe modyfikacje społecznej sytuacji badania (choćby wspomniany poziom penetracji telefonów stacjonarnych i komórkowych).

Na obecnym etapie refleksji *Mixed Mode Survey Design* wydaje się raczej zbiorem racjonalnych postulatów, a nie dopracowaną procedurą badawczą. Nie jest to prosty *trade-off*, lecz raczej *multiple trade-off*. Co więcej, za celowe można uznać rozszerzenie o wymiar „czasowy” „paradygmatu” myślenia o *Mixed Mode Survey Design*. Czy „czas” (długość okresu realizacji) stanowi ograniczenie i jak znaczące w odniesieniu do kosztów badania? Warto postawić dodatkowe pytanie: czy badanie sekwencyjne *Mixed Mode* rozpoczynać od techniki „najszybszej” czy „najtańszej”? Ale, to wszystko zależy . . .

## BIBLIOGRAFIA

- American Association for Public Opinion Research [2011], <http://aapor.org/Content/StandardDefinitions.pdf> [dostęp 5.08.2011].
- Bassler A., Tchaoussoglou C. [2001], *Searching for best CASI practices: experiments for a new data collection method*, 10<sup>th</sup> Worldwide Readership Research Symposium, Venice.
- Biemer P., Caspar R. [1994], *Continuous Quality Improvement for Survey Operations: Some General Principles and Applications*, „Journal of Official Statistics”, vol. 10, no. 3.
- Bowling A. [2005], *Mode of questionnaire administration can have serious effects on data quality*, „Journal of Public Health”, vol. 27.

- Cichomski B., Jerzyński T., Zieliński M. [2009], *Polskie Generalne Sondáže Społeczne: struktura skumulowanych wyników badań 1992–2008*, Warszawa: Wydawnictwo ISS UW.
- Czaja R., Blair J. [2005], *Designing surveys: A guide to decisions and procedures*, Thousand Oaks, CA: Sage Publications.
- de Leeuw E.D. [2008], *Choosing the method of data collection*, [w:] E.D. de Leeuw, J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Leeuw E.D. [2009], *Mixed Mode Surveys – 6<sup>th</sup> ESS Training Course, 30 April–1 May*, Ljubljana, Slovenia.
- de Leeuw E.D., de Heer W. [2002], *Trends in Household Survey Nonresponse: A Longitudinal and International Comparison*, [w:] R.M. Groves, D.A. Dillman, J. L. Eltinge, R.J.A. Little (eds), *Survey Nonresponse*, New York: John Wiley & Sons.
- de Leeuw E.D., Hox J.J., Dillman D.A., [2008a], *The Cornerstones of Survey Research*, [w:] E.D. de Leeuw, J.J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Leeuw E.D., Hox J.J., Dillman D.A., [2008b], *Mixed Mode Surveys: When and Why*, [w:] E.D. de Leeuw, J.J. Hox, D.A. Dillman (eds), *International Handbook of Survey Methodology*, New York–London: Lawrence Erlbaum Associates.
- de Maris A. [1992], *Logit Modeling. Practical Applications*, Thousand Oaks, CA: Sage Publications.
- Dillman D.A. [1978], *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley & Sons.
- Dillman D.A. [2000], *Mail and Internet Surveys: The Tailored Design Method*, New York: John Wiley & Sons.
- Dillman D.A., Smith J., Christian L. [2009], *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, New York: John Wiley & Sons.
- European Social Survey, Round 5. Specification for participating countries [2011], <http://www.europeansocialsurvey.org> [dostęp 5.08.2011].
- Eva G., Widdop S. [2007], *Mixed Mode Data Collection in Europe*, 2<sup>nd</sup> European Survey Research Association Conference, Prague, June.
- Gostkowski Z. [1965], *Badania adekwatności technik badawczych w krajach na drodze rozwoju*, „Przegląd Socjologiczny”, t. XIX/2.
- Gostkowski Z. [1966], *O założeniach i potrzebie empirycznych badań nad technikami i procedurami badawczymi w socjologii*, [w:] Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław: Zakład Narodowy im. Ossolińskich.
- Goyder J.C. [1987], *The Silent Minority*, Boulder, CO: Westview Press.
- Groves R.M. [1989], *Survey Errors and Survey Costs*, New York: John Wiley & Sons.
- Groves R.M. [1991], *Measurement Error Across the Disciplines*, [w:] P.P. Biemer, R.M. Goves, L.E. Lyberg, N. A. Mathiowetz, S. Sudman (eds), *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Groves R.M., Couper M. P. [1998], *Nonresponse in household interview surveys*, New York: John Wiley & Sons.
- Groves R.M., McGonagle K. [2001], *A theory-guided interviewer training protocol regarding survey participation*, “Journal of Official Statistics”, vol. 17.

- Groves R.M., Fowler F.J. Jr., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. [2004], *Survey Methodology*, Hoboken, New Jersey: John Wiley & Sons, Inc.
- Grzeszkiewicz-Radulska K. [2009], *Respondenci niedostępni w badaniach sondażowych*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 12, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Haeder S., Lehnhoff I., Mardian I. [2010], *Mobile Phone Surveys: Empirical Findings from a Research Project*, "ASK. Research and Methods", vol. 19.
- Lutyńska K. [1978], *Ankieterzy i badacze. Z badań nad wpływem ankieterskim*, „Przegląd Socjologiczny”, t. XXX.
- Lutyński J. [1968], *Ankieta i jej rodzaje na tle podziału technik otrzymywania materiałów*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lutyński J. [1972], *Analiza procesu otrzymywania informacji w badaniach z zastosowaniem wywiadu kwestionariuszowego*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 4, Wrocław: Zakład Narodowy im. Ossolińskich.
- Lynn P. [2009], *Mixed or Muddled? Combining Survey Modes in The 21<sup>st</sup> Century*, *Keynote Presentation*, 3<sup>rd</sup> European Survey Research Association Conference, Warsaw, June.
- Martin P. [2010], *Podejście mieszane (Mixed Mode) – możliwości i ograniczenia*, <http://www.konferencja-ewaluacja.pl/2010/materialy.php> [dostęp 5.08.2011].
- Parten M. [1950], *Surveys, Polls and Samples*, New York: Harper & Brothers.
- Roberts C. [2007], *Mixing Modes of Data Collection in Surveys*, 2<sup>nd</sup> European Survey Research Association Conference, Prague, June.
- Roberts C., Eva G., Lynn P., Johnson J. [2010], *Measuring the effect of interview length on response propensity and response quality in a telephone survey – Final report of the ESS CATI experiment*, Centre for Comparative Social Surveys, City University, London.
- Saris W.E. [1991], *Computer-Assisted Interviewing*, Newbury Park: Sage Publications.
- Słomczyński K.M. [1968], *Granice stosowalności ankiety audytoryjnej w środowisku robotniczym*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Zakład Narodowy im. Ossolińskich.
- Stoop I. [2005], *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*, The Hague: Social and Cultural Planning Office of the Netherlands.
- Stoop I., Billiet J., Koch A., Fitzgerald R. [2010], *Improving Survey Response. Lessons learned from the European Social Survey*, New York: John Wiley & Sons.
- Sztabiński F. [1997], *Ankieta pocztowa i wywiad kwestionariuszowy*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński F. [2003], *Logika badacza i logika respondenta. Problem adekwatności narzędzia badawczego*, „ASK. Społeczeństwo, Badania, Metody”, vol. 12.
- Sztabiński F. [2009], *Ankieta internetowa (CAWI): czy rzeczywiście idealna technika?*, [w:] A. Haber, M. Szałaj (red.), *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, Warszawa: PARP.
- Sztabiński F. [2011], *Ocena jakości danych w badaniach surveyowych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński F., Sztabiński P.B. [1994], *Polscy respondenci początku lat 90. Stosunek do badań i jego uwarunkowanie*, „Studia Socjologiczne”, nr 2.

- Sztabiński F., Sztabiński P.B. [1997], *Wartość odpowiedzi w wywiadach telefonicznych ze wspomaganie komputerowym (CATI)*, „ASK. Społeczeństwo, Badania, Metody”, vol. 1–2 (5–6).
- Sztabiński P.B. [1997], *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych?*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński P.B. [1999a], *Wywiad telefoniczny ze wspomaganie komputerowym (CATI) a wywiad kwestionariuszowy. Problem porównywalności wyników*, [w:] H. Domański, K. Lutyńska, A. W. Rostocki (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, Warszawa: Wydawnictwo IFiS PAN.
- Sztabiński P.B. [1999b], *Wywiad telefoniczny ze wspomaganie komputerowym (CATI): co zyskujemy, co tracimy?*, „ASK. Społeczeństwo, Badania, Metody”, vol. 8.
- Sztabiński P.B. [2001], *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywistość idealna technika?*, „ASK. Społeczeństwo, Badania, Metody”, vol. 10.
- Sztabiński P.B., Dyjas-Pokorska A., Żmijewska-Jędrzejczyk T. [2008], *Understanding Refusals*, „ASK. Research and Methods”, vol. 17.
- Sztabiński P.B., Sztabiński F., Przybysz D. [2007], *Are non-respondents similar to respondents? Findings from the ESS-2004 in Poland*. „ASK. Społeczeństwo, Badania, Metody”, vol. 16.
- Weisberg H.F. [2005], *The Total Survey Error Approach. A Guide to the New Science of Survey Research*, Chicago–London: The University of Chicago Press.
- Żmijewska-Jędrzejczyk T. [2004], *Badania internetowe*, [w:] P.B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: Wydawnictwo IFiS PAN.

## Aneks

TABELA I. Schemat badania ESS-CATI r. 3

| Grupa   | Wersja kwestionariusza/moduły  |   |   |   |   |   | Czas      | Próba (%)  |    |                    |
|---------|--|---|---|---|---|---|-----------|------------|----|--------------------|
| Grupa A | Pełna wersja kwestionariusza ESS (wersja A)                                      |   |   |   |   |   | ≈ 60'     | n = 400/40 |    |                    |
|         | moduły:  | A | B | C | D | E |           |            | F  |                    |
| Grupa B | Skrócona wersja kwestionariusza ESS (wersja B)                                   |   |   |   |   |   | ≈ 45'     | n = 200/20 |    |                    |
|         | moduły:  | A | B | C |   | E |           |            | F  |                    |
| Grupa C | Pełna wersja kwestionariusza ESS (wersja C) realizowana w dwóch częściach: a i b |   |   |   |   |   | 2 × ≈ 30' | n = 400/40 |    |                    |
|         | Część a. moduły:   | A | B |   | D |   |           |            | Fx | Moduł Fx (24 pyt.) |
|         | Część b. moduły:   |   |   | C |   | E |           |            | Fy | Moduł Fy (52 pyt.) |

### Moduły:

- A – Media, zaufanie do ludzi
- B – Zainteresowania polityką, zaufanie do instytucji, udział w wyborach, orientacje polityczne i życiowe

- C – Poczucie dobrobytu i wykluczenia społecznego, przynależność religijna, narodowa i etniczna
- D – Kluczowe wydarzenia życiowe, postawy wobec najniższego i najwyższego progu wydarzeń życiowych
- E – Dobrobyt osobisty i społeczny, pomoc innym, odczucia z ubiegłego tygodnia, zadowolenie z życia i pracy
- F (x, y) – Charakterystyki społeczno-demograficzne respondenta i partnera.

Franciszek Sztabiński  
Teresa Żmijewska-Jędrzejczyk

#### MIXED MODE SURVEY DESIGN: MODE EFFECT PROBLEM

##### Abstract

The fieldwork conditions for surveys becomes progressively worse so methods are being sought to reduce the *total survey error* arising from *non-response error*. One solution is the *mixed mode survey design* procedure whereby the same information is obtained from different respondents, using different data collecting modes (e.g. PAPI, CATI, Web survey). However, this procedure may produce the so-called *mode effect*. In this paper, we analyse the mode effect from the following perspectives: (i) wording of questions, (ii) contacting mode, and (iii) mode of communication with the respondent. We assess response rates and compare respondents' answers to the same questions obtained through different techniques.

Our analysis is based on materials collected in research projects employing two modes PAPI (the European Social Survey, Round 3) and CATI (ESS Infrastructure: i3, a study funded under a methodological grant). The findings from our analysis confirm the existence of significant differences related to the *mode effect* (in the case of PAPI and CAPI). Those findings challenge the idea of combining those modes to reduce *non-response error*.

**Key words:** Total Survey Error, Non-Response Error, Mode Effect, Mixed Mode Survey Design, European Social Survey

RAFAŁ BOGUSZEWSKI\*  
NATALIA HIPSZ\*\*  
Centrum Badania Opinii Społecznej

## **OD KWESTIONARIUSZA PAPIEROWEGO DO WSPOMAGANEGO KOMPUTEROWO. ANALIZA PORÓWNAWCZA TECHNIK PAPI I CAPI**

### **Streszczenie**

Dywersyfikacja technik ilościowych w naturalny sposób rodzi pytanie o ich ekwiwalentność. W niniejszym opracowaniu porównano dwie popularne w Polsce procedury badawcze: „tradycyjną” technikę PAPI oraz „nowoczesną” – CAPI. Analizy przeprowadzono w oparciu o doświadczenia Centrum Badania Opinii Społecznej – polskiego ośrodka badawczego, w którym we wrześniu 2008 roku, po dwudziestu sześciu latach badań tradycyjnych, przeprowadzono pierwszy pomiar z wykorzystaniem techniki CAPI. W artykule w pierwszej kolejności opisano zróżnicowanie technik w aspekcie wskaźników realizacyjnych, a następnie zanalizowano wpływ procedur na stosunek respondentów do aktu pomiaru oraz udzielane odpowiedzi.

**Słowa kluczowe:** Centrum Badania Opinii Społecznej, PAPI, CAPI, współczynnik realizacji, trend

### **WPROWADZENIE**

Znamienną cechą badań ilościowych jest ich standaryzacja. Idea standaryzacji obejmuje szereg postulatów odnoszących się zarówno do czynności, jak i środków badawczych. W tym drugim wymiarze za główny wyróżnik technik sondażowych uważa się wykorzystywanie w nich formularzy, których funkcją jest ukierunkowanie pomiaru i ujednoczenie zgromadzonych w jego toku danych empirycznych (por. Lutyński 1983). W początkowej fazie rozwoju badań ilościowych jedynym nośnikiem kwestionariuszy był papier. Wskutek dynamicznego

---

\* r.boguszewski@cbos.pl \*\* n.hipsz@cbos.pl



rozwoju środków komunikowania na odległość oraz przenoszenia na grunt badań sondażowych owoców postępu technologicznego sytuacja ta sukcesywnie ulegała zmianie. W chwili obecnej decydując się na realizację ilościowego projektu badawczego socjolog – przynajmniej teoretycznie – ma do swej dyspozycji całą gamę technik kwestionariuszowych, zróżnicowanych w aspekcie warunków komunikacyjnych i technologicznych realizacji<sup>1</sup>. Coraz dalej idąca dywersyfikacja technik ilościowych w naturalny sposób rodzi pytanie o ich ekwiwalentność. W niniejszym opracowaniu porównano dwie popularne w Polsce procedury badawcze – „tradycyjną” technikę PAPI oraz „nowoczesną” – CAPI.

Zarówno badania PAPI, jak i CAPI mają charakter interaktywny – realizowane są w bezpośredniej, interpersonalnej sytuacji komunikacyjnej przy aktywnym udziale ankietera, który odczytuje respondentowi pytania kwestionariusza, a następnie zaznacza wybrane przez niego odpowiedzi. Zasadniczą różnicą między technikami jest natomiast forma utrwalania danych – o ile w badaniach PAPI odpowiedzi nanoszone są na papierowy kwestionariusz, o tyle w CAPI ankieter wprowadza je do specjalnej aplikacji zainstalowanej w przenośnym komputerze [Schräpler i in. 2006]. Fakt, iż odmienność obydwu podejść ogranicza się do sposobu zapisu materiału empirycznego przesądza o tym, że różnic w ich rezultatach zwykło się upatrywać w większym stopniu po stronie instytucji badawczej i reprezentującego ją ankietera, aniżeli respondenta uwikłanego w obu przypadkach w raczej zbliżony kontekst realizacji [Baker i in. 1991].

Pierwszy europejski projekt badawczy wspomagany komputerowo przeprowadzono w Holandii, w roku 1984 [Schräpler i in. 2006]. Od tego czasu w anglojęzycznej literaturze przedmiotu pojawiło się wiele opracowań o charakterze porównawczym, omawiających różnice między sondażami CAPI i PAPI. Autorzy opracowań wskazują na ogół na profity czerpane z zastosowania bardziej zaawansowanej technologii komputerowej. Powszechnie zwraca się uwagę na oszczędność czasu oraz pieniędzy wynikającą z pomijania etapu transformacji danych z „papieru” do postaci elektronicznej [Nicholls i in. 1997], z drugiej jednak strony zauważa się, iż zastosowanie przez instytucje badawcze procedur CAPI pociąga za sobą znaczne nakłady finansowe związane z zakupem przeno-

---

<sup>1</sup> Badania ilościowe na świecie są prowadzone z wykorzystaniem różnych technik badawczych, np. PAPI (ang. *Paper And Pencil Interviewing*), CAPI (ang. *Computer-Assisted Personal Interviewing*), CASI (ang. *Computer-Assisted Self Interviewing*), ACASI (ang. *Audio Computer-Assisted Self Interviewing*), CAWI (ang. *Computer-Assisted Web Interviewing*), WATI (ang. *Web-Assisted Telephone Interviewing*), CATI (ang. *Computer-Assisted Telephone Interviewing*), VR (ang. *Voice Recognition*), TDE (ang. *Touch-tone Data Entry*), CSAQ (ang. *Computerized Self-Administered Questionnaire*).

śnych komputerów oraz specjalistycznego oprogramowania [Bradburn i in. 1991; Martin i in. 1993]. Okolicznością przemawiającą na korzyść procedur CAPI jest skuteczne eliminowanie przez aplikację błędów ankierskich i koderskich spowodowanych „nietrzymaniem filtrów”, a więc zadawaniem respondentom nieprzeznaczonych dla nich pytań warunkowych, pomijaniem pytań, na które powinni udzielić odpowiedzi, czy wprowadzaniem przypadkowych, niezdefiniowanych w kafeterii wartości [Nicholls i in. 1997; Schröppler i in. 2006]. Także sami ankierzy zwykli postrzegać prowadzenie wywiadów ze wspomaganie laptopa za bardziej nobilitujące od pracy z tradycyjnym narzędziem badawczym [De Leeuw i in. 1995; Martin i in. 1993]. Niektórzy autorzy zauważają ponadto różnice w postawach badanych. Rezultaty eksperymentów metodologicznych dowodzą, iż niekiedy w kwestiach drażliwych odpowiedzi respondentów biorących udział w badaniu CAPI w sposób istotny różnią się od wskazań ankietowanych metodą tradycyjną. Odpowiedzi tych pierwszych okazują się w większym stopniu niezgodne ze społecznie akceptowanymi wartościami (ang. *social desirability bias*) [Bradburn, Sudman 1974; DeMaio 1984; Hyman 1954], a rzadziej ucieczkowe, co wydaje się dowodzić percepcji większej poufności danych zapisywanych w formie elektronicznej [Baker i in. 1995; De Leeuw i in. 1995; Schröppler i in. 2006]. W badaniach ze wspomaganie komputerowym dostrzeżono również bardziej życzliwy stosunek respondentów do aktu pomiaru, przemawiający za słusnością tezy, w myśl której wyposażenie ankiera w sprzęt elektroniczny sprzyja profesjonalizacji jego wizerunku [Baker 1992; Bradburn i in. 1991]. Zbliżony w obydwu technikach, a niekiedy nieznacznie niższy w pomiarach CAPI okazuje się natomiast poziom współczynnika realizacji (ang. *response rate*), a więc – w badaniach wykorzystujących probabilistyczny algorytm doboru respondentów – stosunek liczebności próby zrealizowanej do liczebności wyjściowej wylosowanej z operatu [Schröppler i in. 2006].

Niniejsze opracowanie wykorzystuje doświadczenia Centrum Badania Opinii Społecznej (CBOS) – polskiego ośrodka badawczego, w którym we wrześniu 2008 r., po dwudziestu sześciu latach badań tradycyjnych, przeprowadzono pierwszy pomiar z wykorzystaniem techniki CAPI. Zważywszy na fakt, iż porównanie rezultatów pojedynczych pomiarów nie jest wystarczająco wiarygodne, w tekście prezentowane są dane zagregowane, uwzględniające wyniki pięciu ostatnich pomiarów techniką PAPI (kwiecień 2008 – sierpień 2008) oraz pięciu pierwszych techniką CAPI (wrzesień 2008 – styczeń 2009). Decydując się na porównywanie następujących po sobie pomiarów wzięto pod uwagę to, że zarówno ludzkie opinie, jak i poziom realizacji badań ulegają zmianom w czasie [Dąbała 2007]. Uznano w związku z tym, że czynnik czasu dzielącego pomiary realizowane w analogicz-

nych okresach w kolejnych latach będzie generował większe obciążenie danych, aniżeli uwzględnienie w zestawieniu wyników sąsiadujących ze sobą badań, mimo że tylko jedna z porównywanych grup (badania realizowane z wykorzystaniem techniki PAPI) obejmuje pomiary przeprowadzane w specyficznym dla branży badawczej okresie wakacyjnym. Wszystkie zbiory pochodzą z comiesięcznych badań przeglądowych CBOS („Aktualnie problemy i wydarzenia”). Omawiane sondaże są realizowane na próbach reprezentatywnych dla populacji dorosłych mieszkańców Polski, dla których podstawę doboru od ponad dekady stanowi baza Powszechnego Elektronicznego Systemu Ewidencji Ludności (PESEL). Liczebność wyjściowa każdej z prób w analizowanym okresie obejmowała 2304 osoby.

W artykule w pierwszej kolejności opisano zróżnicowanie technik w aspekcie – szeroko rozumianych – wskaźników realizacyjnych, a następnie zanalizowano wpływ procedur na stosunek respondentów do aktu pomiaru oraz udzielane odpowiedzi.

### WSKAŹNIKI REALIZACYJNE

W badaniach CAPI oprogramowanie komputerowe stanowi jedyny nośnik kwestionariusza wywiadu, czyniąc tym samym notebooka podstawowym narzędziem pracy ankietera. Transformacja z techniki PAPI do CAPI wiązała się z koniecznością zakupu dużej liczby przenośnych komputerów. Wysoki koszt wyposażenia osób realizujących badanie w terenie poskutkowało po wprowadzeniu CAPI redukcją o 21% (ze 176 do 139) przeciętnej liczby ankieterów zaangażowanych w pojedyncze badanie. Przy niezmienionej liczebności wyjściowej próby oznaczało to zwiększenie średniej liczby wywiadów przypadających na jednego ankietera (o 3,5 wywiadu) oraz – w efekcie – wzrost przeciętnej liczby wywiadów zrealizowanych przez jednego ankietera (o 1,5 wywiadu) – patrz tabela 1. Okoliczności te z pewnością nie sprzyjały osiągnięciu satysfakcjonującego poziomu realizacji próby wylosowanej.

TABELA 1. Wskaźniki realizacyjne związane z zaangażowaniem ankieterów

| Wskaźniki realizacyjne:                                     | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| Średnia liczba ankieterów realizujących jedno badanie       | 176                    | 139                      |
| Średnia liczba wywiadów przypadających na jednego ankietera | 13,1                   | 16,6                     |
| Średnia wywiadów zrealizowanych przez jednego ankietera     | 6,2                    | 7,7                      |

Z drugiej jednak strony wskutek zamierzonego zmniejszenia średniej objętości kwestionariusza, w badaniach CAPI ograniczony został przeciętny czas przeprowadzenia pojedynczego wywiadu, a charakterystyczna dla techniki redukcja prac związanych z przygotowaniem zebranych w terenie danych do analiz umożliwiła wydłużenie realizacji terenowej z 6 do 8 dni (patrz tabela 2).

TABELA 2. Wskaźniki realizacyjne związane z narzędziem badawczym

| Wskaźniki realizacyjne:                             | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| Średnia objętość kwestionariusza (liczba zmiennych) | 277                    | 252                      |
| Średni czas realizacji jednego wywiadu w minutach   | 52                     | 42                       |

Koincydencja tych wzajemnie się niwelujących czynników w ostatecznym rozrachunku zaowocowała współczynnikami realizacji o wartościach zbliżonych do rejestrowanych w pomiarach tradycyjnych – choć uśredniony *response rate* wskazuje na nieznacznie mniejszą skuteczność badań CAPI (o 1,5 punktu procentowego), różnica ta jest nieistotna statystycznie. Za porównywalnością realizacji obydwu procedur przemawia ponadto fakt, że rok przed wprowadzeniem CAPI, w analogicznym do rozpatrywanego tu okresie (wrzesień 2007 – styczeń 2008), średni współczynnik osiągał wartość o 1,3 punktu procentowego niższą (44,9%) – patrz tabela 3.

TABELA 3. Wskaźniki realizacyjne związane z próbą

| Wskaźniki realizacyjne:                  | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|--|------------------------|--------------------------|
| Liczebność wylosowana                    | N = 11520              | N = 11520                |
| Liczebność zrealizowana                  | N = 5494               | N = 5327                 |
| Średnia wartość współczynnika realizacji | 47,7%**                | 46,2%**                  |

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

W bezpośredniej relacji z poziomem realizacji pozostają precyzowane przez ankierów powody, dla których wywiady nie zostały przeprowadzone. Mimo że zarówno ogólny ranking przyczyn, jak i rozkład częstości poszczególnych z nich w obydwu porównywanych sytuacjach badawczych jest raczej podobny, odnotowano statystycznie istotne różnice w częstości odmów udziału w badaniu oraz niezastania pod wskazanym adresem nikogo przez cały czas jego realizacji.

TABELA 4. Przyczyny niezrealizowania wywiadów

| Przyczyny:                                      | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| Respondent nieobecny przez cały czas realizacji | 30,3%**                | 31,5%**                  |
| Odmowa udziału w badaniu                        | 22,4%*                 | 26,6%*                   |
| Nikogo nie zastano                              | 18,0%*                 | 13,8%*                   |
| Respondent zmienił miejsce zamieszkania         | 16,7%**                | 15,8%**                  |
| Niesprawność                                    | 3,8%**                 | 3,7%**                   |
| Błędny adres                                    | 1,1%**                 | 1,1%**                   |
| Respondent nie żyje                             | 1,0%**                 | 1,3%**                   |
| Respondent nie spełnia kryteriów                | 0,1%**                 | 0,1%**                   |
| Inne przyczyny                                  | 6,0%**                 | 5,7%**                   |
| Brak danych                                     | 0,6%**                 | 0,4%**                   |

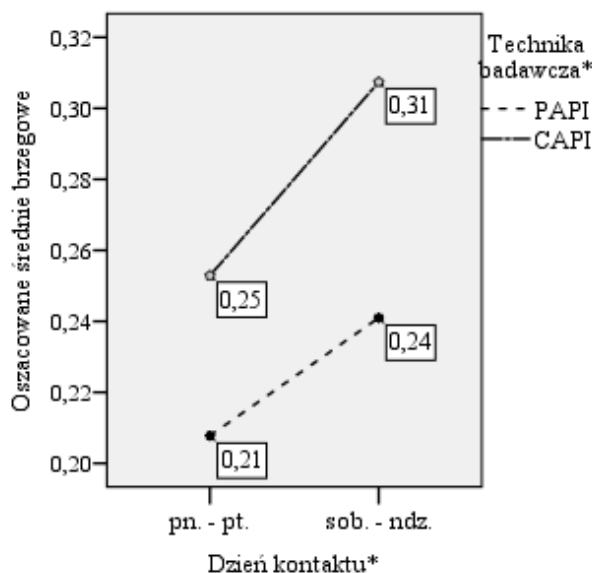
\* Różnica frakcji istotna statystycznie;  $p < 0,05$

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

W rozpatrywanym tu okresie pomiarów tradycyjnych z powodu odmowy nie dochodził do skutku co piąty wywiad niezrealizowany (22,4%), podczas gdy w badaniach wspomaganym komputerowo – co czwarty (26,6%) - patrz tabela 4. Mając na uwadze fakt, iż tej czteropunktowej różnicy nie tłumaczy zmiana proporcji realizacji w dni robocze i weekendy<sup>2</sup> (patrz rysunek 1), może ona wskazywać na nieco mniejszą kooperatywność potencjalnych respondentów w badaniach CAPI. Możliwe, iż widok komputera czy samego dodatkowego „bagażu” przywodzi na myśl niektórych osób niefortunne, negatywne skojarzenia, przez co częściej nie godzą się one na udział w projekcie badawczym (tym bardziej, iż nie są informowane o tym fakcie w poprzedzającym badanie liście zapowiednim).

<sup>2</sup> W badaniach CAPI w wyniku redukcji zatrudnionych ankierów i wydłużenia czasu prac terenowych, zmieniły się proporcje realizacji w dni robocze oraz sobotę i niedzielę. O ile w badaniach PAPI gros wywiadów (58,3%) przeprowadzano w weekend, o tyle w badaniach ze wspomaganym komputerowym realizacja sobotnio-niedzielną obniżyła się (do 45,6%) i rozłożyła na dodatkowe dni przypadające po weekendzie, a więc w tygodniu roboczym. Trzeba jednak zaznaczyć, że większy odsetek odmów w weekendy nie pociąga za sobą niższego współczynnika realizacji: zarówno w badaniach tradycyjnych, jaki i ze wspomaganym przeciętna skuteczność realizacji w soboty i niedziele jest większa niż w dni robocze. W badaniach PAPI różnica ta wynosi 8,8 punktu procentowego (43,0% w dni robocze, 51,8% w weekendy), natomiast w badaniach CAPI – 21,2 punktu procentowego (39,1% w dni robocze, 60,3% w weekendy).

RYSUNEK 1. Frakcja odmów udziału w badaniu w zależności od techniki badawczej i dnia kontaktu

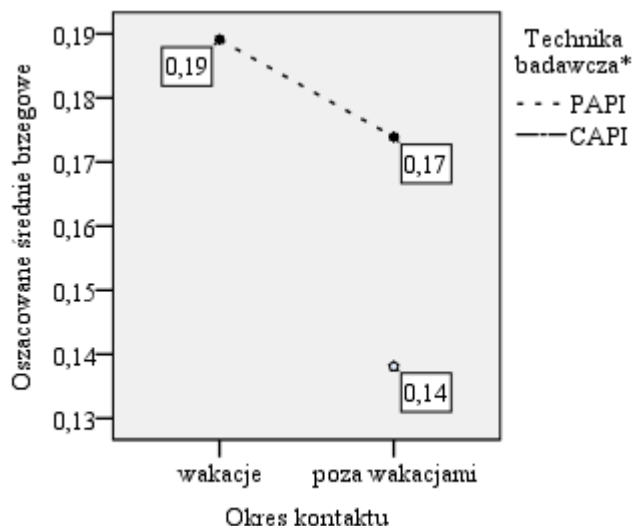


\* Różnica frakcji istotna statystycznie;  $p < 0,05$

Wzrost częstości odmów udziału w badaniu w projektach CAPI odbył się kosztem zmniejszenia liczby wywiadów niezrealizowanych z powodu niezastania nikogo pod wskazanym w próbie adresem. W badaniach tradycyjnych odsetek wywiadów nieprzeprowadzonych ze względu na tego rodzaju trudności wynosił 18,0%, natomiast we wspomaganych komputerowo – 13,8%. Dysproporcja w częstości problemów z zastaniem członków gospodarstwa domowego nie wynika z różnicy w ilości prób kontaktu z wylosowanym respondentem<sup>3</sup>. Nie tłumaczy jej też w pełni fakt realizacji dwóch z pięciu pomiarów PAPI w okresie letnim, kiedy to respondenci są szczególnie trudno uchwytli – różnice w odsetkach utrzymują się także w wynikach badań z okresu niewakacyjnego (patrz rysunek 2). Zważywszy jednak na to, iż trudno upatrywać przyczyny obserwowanych dysproporcji w technologii realizacji wywiadu, należałoby doszukiwać się jej raczej po stronie innych czynników, takich jak specyfika próby badawczej.

<sup>3</sup> W badaniach PAPI przeciętna liczba prób kontaktu, po których rezygnowano z kolejnych wizyt klasyfikując wywiad jako niezrealizowany z powodu niezastania nikogo pod wskazanym adresem przekraczała trzy (3,5). W rozpatrywanym tu okresie projektów CAPI od kolejnych prób kontaktu odstępowano wcześniej – po niespełnieniu trzech zwieńczonych niepowodzeniem wizytach (2,9).

RYSUNEK 2. Frakcja wywiadów niezrealizowanych z powodu niezastania nikogo pod adresem z próby w zależności od techniki badawczej i okresu kontaktu



\* Różnica frakcji istotna statystycznie;  $p < 0,05$

### STOSUNEK RESPONDENTÓW DO BADANIA

Chociaż współczynniki realizacji nie zmieniły się istotnie w związku z przejściem z PAPI na CAPI, pewne różnice obserwujemy w stosunku respondentów do wywiadu – określanym oczywiście na podstawie subiektywnej opinii osób realizujących ankietę. Analiza zagregowanych zbiorów danych pozwala stwierdzić, że w badaniach CAPI podejście ankietowanych do zrealizowanego wywiadu oceniane było przez ankietatorów jako bardziej pozytywne niż w badaniach PAPI, jednocześnie zmniejszył się odsetek odpowiedzi wskazujących na obojętność oraz niechęć wobec wywiadu (patrz tabela 5).

Z pewnością nie bez wpływu na tę korzystną zmianę pozostały wspomniane już wcześniej elementy towarzyszące zmianie metody realizacji badania, a więc mniejsza liczba zmiennych (czyli krótsza ankietka) oraz – związany z tym – krótszy czas realizacji badania, co niewątpliwie sprzyja wskaźnikom realizacji i w mniejszym stopniu „zraża” respondentów do badania.

TABELA 5. Rozkłady odpowiedzi na pytanie do ankietera o ocenę stosunku respondenta do wywiadu

| Stosunek respondentów do wywiadu: | PAPI<br>(IV–VIII 2008)  |        | CAPI<br>(IX 2008–I 2009) |        |
|-----------------------------------|-------------------------|--------|--------------------------|--------|
|                                   | – zdecydowanie życzliwy | 55,9%* | 87,8%*                   | 67,2%* |
| – raczej życzliwy                 | 31,9%*                  | 25,9%* |                          |        |
| – obojętny                        | 9,1%*                   |        | 5,1%*                    |        |
| – raczej niechętny                | 2,5%*                   | 3,1%*  | 1,6%*                    | 1,8%*  |
| – zdecydowanie niechętny          | 0,6%*                   |        | 0,2%*                    |        |

\* Różnica frakcji istotna statystycznie;  $p < 0,05$

Czynnikiem w dalszej kolejności brany pod uwagę jako w pewnym stopniu wpływający na podejście respondentów zarówno do ankietera, jak i do wywiadu jest wizerunek osób realizujących badanie. Jak już wcześniej wspomniano, przyjmuje się, że użycie notebooka w czasie wywiadu sprzyja profesjonalizacji wizerunku ankieterów oraz lepszej ocenie ich prac [Baker 1992; Bradburn i in. 1991].

## ODPOWIEDZI

Na koniec pojawia się chyba najistotniejsze pytanie: czy zmiana narzędzia wpłynęła w jakiś sposób na wyniki badań? Aby to sprawdzić skupimy się na rozkładach odpowiedzi na pytania, po pierwsze, zadawane w badaniach CBOS regularnie co miesiąc, a po drugie, odnoszące się do deklaracji, które zwykle nie ulegają gwałtownym przemianom w krótkim czasie, co umożliwi w miarę rzetelne określenie ewentualnego wpływu metody na rozkłady odpowiedzi. Wskaźniki, na których się skoncentrujemy dotyczą deklaracji udziału w wyborach, samooceny wiary, określenia swoich poglądów politycznych na 7-punktowej skali od lewicy do prawicy oraz oceny sytuacji materialnej własnego gospodarstwa domowego.

Okazuje się, że rozkłady odpowiedzi na pytania dotyczące wszystkich wyżej wymienionych kwestii bez względu na metodę realizacji wywiadu pozostają niemal identyczne wobec wywiadu (patrz tabela 6–9). Obserwowane różnice są zwykle niewielkie i nieistotne statystycznie.

Przeciętne deklaracje udziału w ewentualnych wyborach parlamentarnych, czy to w pięciu ostatnich badaniach realizowanych techniką PAPI, czy też w pięciu pierwszych po zmianie techniki na CAPI, są właściwie tożsame. W obu przypadkach respondenci zamierzający głosować stanowią niespełna trzy piąte badanych, wahający się jedną szóstą, a niezamierzający głosować nieco ponad jedną czwartą. Nie różnią się też deklaracje wiary respondentów. Mniej więcej



co dziesiąty badany określa się jako głęboko wierzący, wierzący stanowią ponad cztery piąte ankietowanych, a niewierzący mniej więcej jedną dwudziestą.

Jeżeli natomiast chodzi o deklarowane poglądy polityczne, pomimo statystycznej istotności, różnice w rozkładach również nie są wielkie i wynikają głównie z dość znaczącego spadku odsetka wskazań „trudno powiedzieć”, który można wiązać z mniej widocznym uwypukleniem tej odpowiedzi w ankietach realizowanych przy pomocy komputerów. Jedynie minimalne różnice w wynikach, również po części związane z mniejszym odsetkiem odpowiedzi wskazujących na niezdecydowanie, dotyczą oceny sytuacji materialnej respondentów i ich rodzin. Generalnie rzecz biorąc można jednak przyjąć, że rozkłady odpowiedzi na pytanie o ocenę sytuacji materialnej, podobnie jak o poglądy polityczne, czy tym bardziej o deklarację wiary i gotowość do uczestnictwa w wyborach – bez względu na metodę realizacji badania – pozostają zbliżone i nie ma podstaw by przypuszczać, że zmiana techniki w istotny sposób wpłynęła na porównywalność wyników.

TABELA 6. Rozkłady odpowiedzi na pytanie o zamiar uczestnictwa w ewentualnych wyborach parlamentarnych

| Gdyby w najbliższą niedzielę odbywały się wybory do Sejmu i Senatu, to czy wziął(ę)by Pan(i) w nich udział? | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| – na pewno wziął(ę)bym w nich udział  | 56,8%**                | 56,8%**                  |
| – nie wiem, czy wziął(ę)bym w nich udział   | 16,9%**                | 16,2%**                  |
| – na pewno nie wziął(ę)bym w nich udziału   | 26,3%**                | 27,0%**                  |

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

TABELA 7. Rozkłady odpowiedzi na pytanie o wiarę

| Niezależnie od udziału w praktykach religijnych, czy uważa się Pan(i) za osobę: | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| – głęboko wierzącą  | 10,2%**                | 10,9%**                  |
| – wierzącą  | 85,2%**                | 83,8%**                  |
| – raczej niewierzącą  | 2,8%**                 | 3,4%**                   |
| – całkowicie niewierzącą  | 1,7%**                 | 1,9%**                   |

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

TABELA 8. Rozkłady odpowiedzi na pytanie o poglądy polityczne

| Proszę wskazać, który z punktów najlepiej odpowiada Pana(i) poglądom politycznym: | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|---|------------------------|--------------------------|
| – lewica (punkty 1–3)   | 12,7%*                 | 14,2%*                   |
| – centrum (punkt 4)   | 28,3%*                 | 32,4%*                   |
| – prawica (punkty 5–7)  | 32,9%**                | 32,5%**                  |
| – trudno powiedzieć   | 26,1%*                 | 20,9%*                   |

\* Różnica frakcji istotna statystycznie;  $p < 0,05$ \*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$ 

TABELA 9. Rozkłady odpowiedzi na pytanie o jakość życia

| Czy obecnie Panu(i) i Pana(i) rodzinie żyje się: | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008 – I 2009) |
|--|------------------------|----------------------------|
| – bardzo dobrze                                  | 3,4%*                  | 4,3%*                      |
| – dobrze   | 36,1%**                | 36,7%**                    |
| – ani dobrze, ani źle                            | 44,9%**                | 44,4%**                    |
| – źle  | 12,7%**                | 11,7%**                    |
| – bardzo źle                                     | 2,6%**                 | 2,8%**                     |
| – trudno powiedzieć                              | 0,3%*                  | 0,1%*                      |

\* Różnica frakcji istotna statystycznie;  $p < 0,05$ \*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$ 

Innym obszarem, w którym badacze odnotowują różnice między wynikami projektów PAPI i CAPI, są odpowiedzi na pytania drażliwe, a więc odnoszące się do regulowanych społecznie norm, których nie wypada nie respektować, bądź poruszające kwestie krępujące, w odczuciu respondenta nazbyt intymne. Obserwowane niekiedy rozbieżności w rezultatach badań tradycyjnych i wspomnianych komputerowo dotyczą częstości wskazań ucieczkowych – przede wszystkim odmów udzielenia odpowiedzi na pytanie.

Informacją osobistą, o której podanie proszeni są wszyscy respondenci CBOS, jest wysokość miesięcznych dochodów: indywidualnych oraz przypadających na jednego członka gospodarstwa domowego. Większą powściągliwością w ujawnianiu wysokości dochodów osobistych wykazywali się ankietowani techniką CAPI – rzadziej w odpowiedzi na pytanie podawali oni kwotę (60,9% do 65,5% w badaniach PAPI), natomiast częściej nie potrafili wskazać konkretnej sumy (7,9% do 6,9% w badaniach PAPI) lub odmawiali jej podania (18,4% do 15,5% w badaniach PAPI) – patrz tabela 10. W tym samym okresie inaczej rozkładały się

odpowiedzi na pytanie o wysokość dochodów *per capita*. Mniej wartościowych danych uzyskano tu w projektach realizowanych w sposób tradycyjny (61,5% do 67,9% w badaniach CAPI), w których częstsze okazały się też odpowiedzi niejednoznaczne (21,3% do 15,9% w badaniach CAPI) oraz odmowy udzielenia takiej informacji (16,6% do 15,4% w badaniach CAPI) – choć w ostatnim przypadku zaobserwowana różnica nie jest istotna statystycznie. Mając więc na uwadze odmienny kierunek odnotowanych różnic, ich wystąpienia nie można przypisać zastosowanym technikom badawczym (patrz tabela 11).

TABELA 10. Rozkłady odpowiedzi na pytanie o wysokość miesięcznych dochodów osobistych

| Sytuacje:  | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008–I 2009) |
|--|------------------------|--------------------------|
| Respondent podaje konkretną kwotę                  | 65,5%*                 | 60,9%*                   |
| Respondent deklaruje brak dochodów                 | 12,1%**                | 12,8%**                  |
| Respondent nie potrafi określić wysokości dochodów | 6,9%*                  | 7,9%*                    |
| Respondent odmawia podania wysokości dochodów      | 15,5%*                 | 18,4%*                   |

\* Różnica frakcji istotna statystycznie;  $p < 0,05$

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

TABELA 11. Rozkłady odpowiedzi na pytanie o wysokość miesięcznych dochodów *per capita*

| Sytuacje:  | PAPI<br>(IV–VIII 2008) | CAPI<br>(IX 2008 – I 2009) |
|--|------------------------|----------------------------|
| Respondent podaje konkretną kwotę                  | 61,5%*                 | 67,9%*                     |
| Respondent deklaruje brak dochodów                 | 0,6%*                  | 0,8%*                      |
| Respondent nie potrafi określić wysokości dochodów | 21,3%*                 | 15,9%*                     |
| Respondent odmawia podania wysokości dochodów      | 16,6%**                | 15,4%**                    |

\* Różnica frakcji istotna statystycznie;  $p < 0,05$

\*\* Różnica frakcji nieistotna statystycznie;  $p > 0,05$

Technologia realizacji wywiadów przy użyciu komputera, jak już wspomniano, stwarza wiele różnych, zwykle niedostępnych w PAPI, dodatkowych możliwości w zakresie konstrukcji narzędzia czy sposobu zadawania pytań, takich jak stosowanie rotacji, wizualizacji graficznej, efektów dźwiękowych, ukrywania odpowiedzi typu „trudno powiedzieć” itp. Wszystkie te „udogodnienia” wprawdzie mogą w różny sposób wpływać na jakość uzyskiwanych odpowiedzi,

zwykle jednak stosowane są w celu pełniejszego przedstawienia problemu, dzięki czemu egzekwują – jak można przypuszczać – odpowiedzi bardziej precyzyjne, a zarazem w większym stopniu odpowiadające rzeczywistości.

Co jednak najważniejsze, technika realizacji badań przy użyciu komputera nie przekreśla możliwości dalszego śledzenia trendów zapoczątkowanych w PAPI poprzez stosowanie identycznych pytań i zadawanie ich w niezmięnionej formie. Nie wszystkie możliwości jakie daje realizacja badania techniką CAPI muszą być w danej sytuacji wykorzystane, zwłaszcza jeżeli zależy nam na zachowaniu porównywalności wyników. Zamiast całkowitego ukrycia odpowiedzi „trudno powiedzieć”, które w ankietach papierowych były widoczne, można umieścić je w kwestionariuszu elektronicznym stosując odmienną (np. mniej widoczną) czcionkę, co – jak się wydaje – dobrze funkcjonuje w porównywalnych pytaniach zadawanych przez CBOS.

Okazuje się jednak, że nawet bardziej radykalne zmiany zastosowane w pytaniach wcześniej zadawanych w kwestionariuszu papierowym nie muszą zaburzać porównywalności wyników. Przyjmuje się na przykład, iż zastosowanie rotacji w pytaniach z długą listą czynników, do których ma się ustosunkować respondent, sprawia, że wyniki będą bardziej wiarygodne, gdyż w związku ze znużeniem narastającym wraz z odczytywanymi podpunktami, odpowiedzi na kwestie z końca listy mogą być przypadkowe.

Pytanie o stosunek Polaków do długiej listy różnych narodów, zadawane co roku przez CBOS, nie wskazuje na taką prawidłowość. Wprawdzie odpowiedzi uzyskane na to pytanie w badaniu zrealizowanym techniką CAPI z zastosowaniem rotacji, dość istotnie różnią się od tych uzyskanych w poprzednim pomiarze realizowanym metodą tradycyjną bez rotacji (państwa były odczytywane w kolejności alfabetycznej), wynika to jednak raczej ze zmiany ogólnego nastawienia Polaków do innych narodów niż z samego zastosowania rotacji. Wskazuje na to fakt, że zmiany sympatii oraz niechęci do państw z końca alfabetycznej listy stosowanej w badaniach PAPI nie różnią się szczególnie, jeżeli chodzi o kierunek i zakres w porównaniu do zmian, jakie zaszły w stosunku do innych państw, np. tych z początku listy (patrz tabela 12).

Prawidłowości takiej nie zanotowano również po zastosowaniu rotacji w comiesięcznym badaniu zaufania do około dwudziestu polityków, którzy w wywiadach realizowanych przy pomocy papierowego kwestionariusza odczytywani byli w kolejności alfabetycznej, natomiast w badaniach CAPI kolejność odczytu nazwisk polityków jest losowa.

TABELA 12. Zmiany sympatii oraz niechęci do poszczególnych narodów

| Narody            | Zmiany sympatii do narodów |                  |          | Zmiany niechęci do narodów |                  |          |
|-------------------|----------------------------|------------------|----------|----------------------------|------------------|----------|
|                   | PAPI<br>VIII 2007          | CAPI<br>XII 2008 | różnica: | PAPI<br>VIII 2007          | CAPI<br>XII 2008 | różnica: |
|                   | w procentach               |                  |          |                            |                  |          |
| Amerykanie        | 44                         | 47               | +3       | 21                         | 18               | -3       |
| Anglicy           | 53                         | 51               | -2       | 14                         | 15               | +1       |
| Arabowie          | 12                         | 21               | +9       | 55                         | 49               | -6       |
| Austriacy         | 36                         | 43               | +7       | 24                         | 18               | -6       |
| Belgowie          | 40                         | 42               | +2       | 16                         | 13               | -3       |
| Białorusini       | 27                         | 34               | +7       | 34                         | 26               | -8       |
| Bułgarzy          | 29                         | 35               | +6       | 28                         | 20               | -8       |
| Chińczycy         | 22                         | 31               | +9       | 38                         | 30               | -8       |
| Czesi             | 53                         | 53               | 0        | 15                         | 12               | -3       |
| Duńczycy          | 41                         | 45               | +4       | 17                         | 14               | -3       |
| Estończycy        | 30                         | 33               | +3       | 22                         | 21               | -1       |
| Francuzi          | 52                         | 49               | -3       | 14                         | 15               | +1       |
| Grecy             | 49                         | 46               | -3       | 12                         | 14               | +2       |
| Gruzini           | 27                         | 35               | +8       | 28                         | 25               | -3       |
| Hiszpanie         | 49                         | 52               | +3       | 13                         | 10               | -3       |
| Holendrzy         | 48                         | 46               | -2       | 12                         | 12               | 0        |
| Irlandczycy       | 54                         | 47               | -7       | 11                         | 13               | +2       |
| Japończycy        | 34                         | 42               | +8       | 24                         | 19               | -5       |
| Litwini           | 38                         | 41               | +3       | 22                         | 17               | -5       |
| Łotysze           | 30                         | 33               | +3       | 25                         | 21               | -4       |
| Niemcy            | 30                         | 38               | +8       | 39                         | 32               | -7       |
| Ormianie          | 20                         | 27               | +7       | 33                         | 27               | -6       |
| Romowie (Cyganie) | 14                         | 21               | +7       | 59                         | 51               | -8       |
| Rosjanie          | 24                         | 30               | +6       | 46                         | 41               | -5       |
| Rumuni            | 16                         | 25               | +9       | 51                         | 40               | -11      |
| Serbowie          | 19                         | 29               | +10      | 40                         | 27               | -13      |
| Słowacy           | 48                         | 51               | +3       | 15                         | 13               | -2       |
| Szwedzi           | 44                         | 45               | +1       | 15                         | 12               | -3       |
| Turcy             | 18                         | 28               | +10      | 46                         | 37               | -9       |
| Ukraińcy          | 25                         | 34               | +9       | 39                         | 31               | -8       |
| Węgrzy            | 45                         | 48               | +3       | 16                         | 13               | -3       |
| Wietnamczycy      | 18                         | 28               | +10      | 41                         | 31               | -10      |
| Włosi             | 51                         | 54               | +3       | 13                         | 12               | -1       |
| Żydzi             | 23                         | 34               | +11      | 40                         | 32               | -8       |

## PODSUMOWANIE

Zmiana techniki realizacyjnej w CBOS nie przyniosła znaczących różnic ani w poziomie realizacji, ani w wynikach badań – świadczą o tym zarówno wskaźniki realizacyjne, jak i rozkłady odpowiedzi na pytania o ewentualny udział w wyborach oraz obserwacje w miarę stałych nastawień, takich jak np. autodeklaracje wiary czy poglądów politycznych, niezmieniających się w czasie tak szybko, jak chociażby opinie.

Jedyną zauważalną zmianą jest spadek udziału odpowiedzi „trudno powiedzieć” w badaniach CAPI, co przypuszczalnie spowodowane jest częściowym lub całkowitym ukryciem możliwości wyboru tego typu odpowiedzi w ankietach realizowanych przy użyciu komputera, podczas gdy w kwestionariuszach papierowych były one w pełni widoczne. Wydaje się to o tyle warte zasygnalizowania, że – jak wynika z dodatkowych pytań, na które odpowiada ankieter CBOS realizujący wywiad CAPI – niemal połowa respondentów śledzi poprawność zapisu udzielanych odpowiedzi (nie ma podstaw, by sądzić, że w badaniach PAPI odsetek ten był znacząco różny). Z drugiej jednak strony należy mieć na uwadze, że ukrycie wariantu „trudno powiedzieć” w badaniach realizowanych techniką komputerową zwykle tylko w niewielkim stopniu ograniczyło wybór tej odpowiedzi. Różnice te najczęściej nie są aż tak istotne, by znacząco modyfikowały rozkłady odpowiedzi na pytania i zaburzały ich porównywalność. Co więcej, nawet bardziej radykalne zmiany w sposobie zadawania pytań związane z wykorzystaniem dodatkowych możliwości, jakie stwarza technika CAPI (np. rotacja pytań lub odpowiedzi), zdają się również nie wpływać znacząco na jakość uzyskiwanych wyników.

Jak to jednak zwykle bywa, każda ze stosowanych metod w realizacji badań ma swoje mocniejsze i słabsze strony. Również w konfrontacji PAPI – CAPI da się wskazać atuty zarówno jednej, jak i drugiej techniki. Na korzyść badań realizowanych za pomocą papierowych kwestionariuszy przemawia to, iż pomimo spadku cen urządzeń elektronicznych koszt komputerów przenośnych, ich utrzymania, a także koszty oprogramowania ciągle przewyższają te, które są generowane przy realizacji badań tradycyjnych, a więc cenę papieru, wydruku oraz koszty wprowadzania i czyszczenia danych.

Zaletą badań PAPI jest niewątpliwie mniej obciążająca możliwość zwiększenia wydajności sieci ankieterskiej poprzez zatrudnienie i przeszkolenie większej liczby osób. Badania CAPI oprócz tego, że wymagają od ankieterów sprawnego posługiwania się komputerem (co stanowi pewne ograniczenie), stwarzają konieczność wyposażenia każdej osoby realizującej badanie w notebooka z odpo-

wiednim oprogramowaniem, co przy ograniczonych zasobach firmy może być kłopotliwe.

Pojawia się również kwestia badań szczególnie drażliwych, które wymagają samodzielnego wypełnienia ankiety przez respondenta. Jak się wydaje, w tym wypadku również lepiej „radzi sobie” papier, co łączy się z umiejętnościami związanymi z obsługą komputera przez respondentów. Problem obsługi komputera, a zwłaszcza sprawności w pisaniu na klawiaturze – tym razem również w odniesieniu do ankietów – wiąże się także, jak wynika z pobieżnych obserwacji wyników badań realizowanych w CBOS, z uzyskiwaniem w badaniach CAPI mniej rozbudowanych, a przez to bardziej ogólnikowych odpowiedzi na pytania otwarte.

Pomimo jednak istotnych zalet przypisywanych tradycyjnej formie realizacji wywiadów, przewaga nowocześniejszej techniki – wspomaganej komputerowo – wydaje się bezdyskusyjna. Za atut badań CAPI – oprócz korzyści wynikających z tego, że ankietę wyposażony w laptopa wygląda bardziej profesjonalnie i tak też jest oceniana jego praca – można uznać fakt, że dzięki odpowiedniemu oprogramowaniu udaje się właściwie wyeliminować braki danych, a automatyczne reguły przejścia pozwalają na wyzbycie się błędów ankieterskich związanych z zadawaniem niewłaściwych pytań oraz omijaniem tych, które powinny zostać zadane.

Niewątpliwą zaletą badań wspomaganych komputerowo jest również to, że pozwalają one na zastosowanie podczas wywiadu różnych elementów niewerbalnych, co wydaje się szczególnie istotne w badaniach o charakterze marketingowym, np. w przypadku testowania rozpoznawalności marki czy oceny jej wizerunku. W wywiadach realizowanych przy użyciu komputera możliwe jest ponadto wykorzystanie różnego rodzaju elementów dźwiękowych i wizualnych, co z kolei ułatwia prowadzenie testów reklam czy produktów.

Komputery stwarzają ponadto zdecydowanie więcej możliwości w zakresie kontroli badań, które nie są dostępne w sondażach realizowanych tradycyjną metodą. Chodzi tutaj m.in. o monitorowanie czasu trwania całego wywiadu oraz jego składowych elementów czy chociażby możliwość odsłuchu przeprowadzonego wywiadu. Już sama świadomość istnienia dodatkowych sposobów weryfikacji pracy ankietów sprzyja, jak się wydaje, wyższej jakości realizowanych przez nich badań.

Najważniejszym wnioskiem z naszych analiz jest jednak to, że obie techniki – PAPI i CAPI – pozwalają zachować porównywalność, a przejście z jednej na drugą nie oznacza braku możliwości kontynuacji pomiaru trendów. W tym względzie ekwiwalentność obu technik nie powinna budzić większych zastrzeżeń.

Zupełnie inną kwestią jest jakość badań i ich wyników, dla której – niezależnie od stosowanego podejścia badawczego – to czynnik ludzki, a więc m.in. rzetelność i uczciwość ankieterów, pozostaje kluczowy.

*Podziękowania za współpracę przy pisaniu artykułu dla Artemis Bellos, Amelii Bieleckiej-Kryski i dr hab. Mirosławy Grabowskiej.*

## BIBLIOGRAFIA

- Baker R.P. [1992], *New technology in survey research: Computer-assisted personal interviewing (CAPI)*, "Social Science Computer Review", vol. 10.
- Baker R.P., Bradburn N.M., Johnson R. [1995], *Computer-assisted personal interviewing: An experimental evaluation of data quality and survey costs*, "Journal of Official Statistics", vol. 11.
- Baker R.P., Bradburn N.M., Frankel M.R., Pergamit M.R. [1991], *A comparison of computer-assisted personal interviews (CAPI) with paper-and-pencil (PAPI) interviews in the National Longitudinal Study of Youth*, „NLS Discussion Paper AAPOR Conference”, Phoenix.
- Bradburn N. M., Sudman S. [1974], *Response effects in surveys: A review and synthesis*, Chicago: Aldine.
- Bradburn N.M., Frankel M., Hunt E., Ingles J., Pergamit M.R., Schoua-Glusberg A., Wojcik M. [1991], *A comparison of computer-assisted personal interviews (CAPI) with personal interviews in the National Longitudinal Survey of Labor Market Behavior-Youth Cohort*, „NLS Discussion Paper AAPOR Conference”, Phoenix.
- Dąbala W. [2007], *Próby CBOS oraz szacowanie na ich podstawie parametrów populacji w badaniach sondażowych po roku 2000*, „Przegląd Socjologiczny”, t. LVI/1.
- De Leeuw E., Hox J., Snijders G. [1995], *The effect of computer-assisted interviewing on data quality*, "International Journal of Market Research", vol. 37(4).
- DeMaio T.J. [1984], *Social desirability and survey measurement: A review*, [w:] Ch. E. Turner, E. Martin (eds), *Surveying subjective phenomena*, vol. 2, New York Russell Sage.
- Hyman H. [1954], *Interviewing in social research*, Chicago: University of Chicago Press.
- Lutyński J. [1983], *Wywiad kwestionariuszowy jako technika otrzymywania danych*, [w:] K. Lutyńska, A.P. Wejland (red.), *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Martin J., O’Muircheartaigh C., Curtice J. [1993], *The use of CAPI for attitude surveys: An experimental comparison with traditional methods*, "Journal of Official Statistics", vol. 9.
- Nicholls W.L., Baker R.P., Martin J. [1997], *The effect of new data collection technologies on survey data quality*, [w:] L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, D. Trewin (eds), *Survey measurement and process quality*, New York: Wiley-Interscience.
- Schräpler J.P., Schupp J., Wagner G.G. [2006], *Changing from PAPI to CAPI: A Longitudinal study of mode-effects based on an experimental design*, „Discussion Paper”, no. 593, Berlin.



*Rafał Boguszewski, Natalia Hipsz*

**FROM THE PAPERQUESTIONNAIRE TO THE COMPUTER-ASSISTED  
ONE THE COMPARISON OF PAPI AND CAPI**

Abstract

The diversification of quantitative techniques naturally raises the question of their equivalence. The paper makes a comparison of two standardized research techniques popular in Poland: “the traditional one” – PAPI and “the modern one” – CAPI. It is based on the experience of Public Opinion Research Center – Polish research center, where in September 2008, after twenty-six years of traditional researches, the first measurement was carried out using CAPI technology. The article first describes the diversity of techniques in terms of realization indicators, then analyzes the impact of procedures on the respondents’ attitude to the act of the measurement and answers given to survey questions.

**Key words:** Public Opinion Research Center, PAPI, CAPI, response rate, trend

KATARZYNA GRZESZKIEWICZ-RADULSKA\*  
Uniwersytet Łódzki

## JAK W POLSCE ROZWIJA SIĘ ZJAWISKO NIEDOSTĘPNOŚCI RESPONDENTÓW? ANALIZA NA PRZYKŁADZIE BADAŃ CBOS 1993–2011

### Streszczenie

Celem artykułu jest przedstawienie i objaśnienie trendów niedostępności respondentów w oparciu o dane z lat 1993–2011 dla sondażu „Aktualne problemy i wydarzenia”, realizowanego techniką wywiadu osobistego przez Centrum Badania Opinii Społecznej. W pracy autorka posługuje się miarami poziomu realizacji przyjętymi przez AAPOR. Wyniki wskazują, że na przestrzeni badanego okresu odsetek odpowiedzi (RR2) spadł o około 40 punktów procentowych, głównie za sprawą wzrostu nieobecności oraz odmów. W analizach wielozmiennowych ustalono, że na kształt krzywej kontaktów (CON1) niekorzystnie wpłynęła przede wszystkim zmiana sposobu doboru próby (przejście z prób gospodarstw domowych na próby imienne) oraz większe obciążenie ankietera pracą. Zmiana sposobu doboru respondentów miała jednak korzystny wpływ na poziom kooperacji (COOP4). Czynniki niesprzyjające partycypacji w sondażach to większe obciążenie ankietera pracą, a także negatywne nastroje społeczne i poprawa sytuacji ekonomicznej w kraju.

**Słowa kluczowe:** niedostępność respondentów, współczynnik kontaktów, współczynnik kooperacji, próba imienna, próba gospodarstw domowych, PAPI, CAPI

### WPROWADZENIE

Zainteresowanie metodologów problemem ubytków w próbie bierze się z niepokoju o prawomocność osiągniętych rezultatów i możliwość generalizacji wniosków, jaką w założeniu daje dobór losowy. Wraz z obserwacją spadających

---

\* katarzyna.radulska@interia.pl katarzyna.radulska@gmail.com

odsetków realizacji, niepokój ten bardzo przybrał na sile, a zagadnienie niedostępności stało się głównym przedmiotem badań metodologii sondażowej.

Celem artykułu jest przedstawienie i objaśnienie trendów niedostępności respondentów w okresie przypadającym na czas trwania III RP. Dla zilustrowania zachodzących zmian (oddzielnie w zbiorze odmów i nieobecności) przeprowadzono analizy na podstawie unikatowego materiału, jakim są dane Centrum Badania Opinii Społecznej (CBOS)<sup>1</sup> z przebiegu realizacji sondaży „Aktualne problemy i wydarzenia” (APiW) – najdłuższego, o najczęściej powtarzanych pomiarach, badania *trackingowego* w Polsce. Dostarczona dokumentacja rozwoju zjawiska niedostępności opiera się na danych dla kraju oraz klas miejscowości. W pracy podejmuję także ostrożną próbę udzielenia odpowiedzi na pytanie o determinanty obserwowanych tendencji, uwzględniając dodatkowo czynniki będące pod kontrolą badacza.

### BAZA MATERIAŁOWA

W prowadzonych przez CBOS sondażowych badaniach opinii publicznej „Aktualne problemy i wydarzenia” wykorzystywaną techniką otrzymywania materiałów jest indywidualny wywiad *face-to-face*, przy czym we wrześniu 2008 r. zrezygnowano z jego tradycyjnej wersji z kwestionariuszem w papierowej formie (PAPI) na rzecz wersji wspomaganą komputerowo (CAPI). Sondaż ten przeprowadzany jest raz na miesiąc w przeciągu kilku dni obejmujących weekend. W badaniu uczestniczy duża liczba ankierów, którzy mają przydzielane wiązki blisko zlokalizowanych adresów, co wynika z klustrowanego charakteru próby. W początkowym okresie wielkość wiązki wynosiła 6, a w tej chwili 8 adresów. Wizytę ankiera poprzedza się wysłanym pocztą listem zapowiednim.

Materiał obejmuje dane pochodzące z dwóch rodzajów prób losowych. W okresie od roku 1993 do sierpnia 2003 r. stosowano próby adresowe<sup>2</sup>, a od września 2003 r. wykorzystuje się próby imienne<sup>3</sup>. W przypadku jednego i drugiego rodzaju doboru, badaną populacją są dorośli mieszkańcy Polski. Na przestrzeni badanego okresu zmieniała się także wielkość prób założonych – z uwagi na

<sup>1</sup> Autorka artykułu serdecznie dziękuje Centrum Badania Opinii Społecznej za udostępnienie materiałów badawczych.

<sup>2</sup> Ogólnopolskie próby gospodarstw domowych, wykorzystujące operat GUS, o trójstopniowym schemacie losowania: rejon statystyczny; mieszkanie; dorosła osoba zamieszkała na stałe, losowana przez ankiera metodą Kisha.

<sup>3</sup> Przy wykorzystaniu operatu PESEL, w którym jednostką jest osoba dorosła, posiadająca w Polsce meldunek lub ubezpieczenie zdrowotne, albo będąca obywatelem polskim zamieszkałym zagranicą i ubiegającym się o wydanie paszportu.

spadające odsetki realizacji liczebność próby była systematycznie zwiększana (w 1993 r. wynosiła 1500, a w 2011 r. – 2384). Problem ten, zaostrzający się szczególnie w wielkich miastach, był dodatkowo powodem stosowania prób nieproporcjonalnych i zwiększania w nich nadreprezentacji jednostek miejskich [Dąbała 2007].

Materiał obejmuje lata 1993–2011, jednak z uwagi na trudność z uzyskaniem dostępu do całości zbioru – z wyłączeniem roczników 1994, 1996, 1999. Z kolei dla roku 2011 dysponuję tylko dwoma pomiarami ze stycznia i lutego. W sumie analiza obejmuje wyniki 171 sondaży, z czego 82 odpowiada próbom adresowym, a 89 imiennym. Podane wielkości odpowiadają materiałom po dokonaniu selekcji, która polegała na wyłączeniu z analiz tych sondaży, przy których realizacji odstąpiono od standardowej, losowej procedury doboru jednostek (np. uruchamianie doboru *random route* od adresu, w którym nie udało się zrealizować wywiadu)<sup>4</sup>.

Podstawę źródłową stanowią relacje ankierów opisujące przebieg ich wizyt pod wskazanymi adresami, utrwalane za pomocą kwestionariusza nazwanego kartą realizacji badania. Narzędzie to podlegało w czasie różnym modyfikacjom, a jedną z najważniejszych było odstępnie w maju 2003 r. od żądania opisu trzech wizyt (dwóch pierwszych i ostatniej) na rzecz jednego, końcowego zapisu<sup>5</sup>. Materiał przekazany przez CBOS miał postać elektronicznych baz danych.

### MIARY POZIOMU REALIZACJI

Prezentacja trendów w zakresie niedostępności respondentów wymaga przyjęcia dokładnie zdefiniowanych miar poziomu realizacji. W Polsce nie funkcjonują jeszcze w tym względzie obowiązujące standardy, a potrzeba ujednoczenia sposobu kodowania ubytków w próbie i operowania wspólnymi współczynnikami realizacji jest rzadko zgłaszana. Tymczasem korzyści wynikające z posiadania takich standardów to – na polu nauki – porównywalność wyników sprzyjająca prowadzeniu systematycznych badań nad niedostępnością respondentów, a na rynku – ograniczenie problemu nieuczciwej konkurencji wśród firm badawczych, „osiągających” wysokie odsetki realizacji, wyliczane według własnych sposobów. Zgłaszany postulat stosowania wspólnych definicji został ostatecznie wcielony

<sup>4</sup> Z przekazanych materiałów za lata 1993 – sierpień 2003 wyłączono 8 sondaży, a za okres wrzesień 2003 – luty 2011 wyłączono 1 sondaż. Pełną charakterystykę materiałów z lat 1993–2003, wraz z informacjami o wprowadzonych poprawkach, czytelnik znajdzie w: [Grzeszkiewicz-Radulska 2009: 37–39, 351–362].

<sup>5</sup> Transformacja wartości zmiennych odpowiadających trzem wizytom, która prowadzi do uzyskania jednego końcowego kodu, została przeprowadzona zgodnie z wytycznymi AAPOR.

w życie przez badaczy należących do The American Association for Public Opinion Research (AAPOR), którzy w 1998 r. wydali pierwszą wersję *Standard Definitions*. Opracowanie to ma w tej chwili już 7 edycji, a jego rozszerzanie wynika z potrzeby dostarczenia narzędzi do badania różnych typów niedostępności dla możliwie wszystkich technik gromadzenia materiałów i metod doboru próby [AAPOR 2011]. Jakkolwiek propozycja AAPOR nie jest jeszcze pełna, upowszechniła się w wielu środowiskach badawczych, stanowiąc bezpośrednie źródło wytycznych albo punkt wyjścia do opracowania analogicznych zasad dla sondaży o specyficznych cechach [Slater, Christensen 2002], często nietypowych dla badań amerykańskich. Takie inicjatywy obserwuje się już wśród badaczy europejskich – tu na uwagę zasługuje opracowanie angielskie, w którym podjęto m.in. próbę dostosowania standardów AAPOR do badań realizowanych techniką wywiadu *face-to-face* na próbach imiennych oraz rozwiązań niestosowanych w USA [Lynn i in. 2001]. Innym wskaźnikiem ważności ustaleń terminologicznych AAPOR jest ich akceptacja w najwyższej notowanych czasopismach naukowych, takich jak „Public Opinion Quarterly” czy „International Journal of Public Opinion Research” [AAPOR 2011: 3].

Zaproponowany przez AAPOR sposób kodowania przypadków wypadających z próby i wyliczania współczynników realizacji opiera się na głównej zasadzie, która mówi, że należy oddzielić błędy pokrycia od błędów niedostępności, a status respondenta niedostępnego przysługuje jednostkom, z którymi nie udało się zrealizować badania, pod warunkiem, że należą do badanej populacji. Zgodnie z tą zasadą, przy uwzględnianiu realiów pracy w terenie, każdą wylosowaną do próby jednostkę należy przyporządkować do jednej z czterech głównych kategorii:

- wywiad zrealizowany;
  - wywiad niezrealizowany, jednostka należy do badanej populacji;
  - wywiad niezrealizowany, nie wiadomo czy jednostka należy do badanej populacji;
  - wywiad niezrealizowany, jednostka nie należy do badanej populacji
- [AAPOR 2011: 7–8.

Ze względu na to, że każda technika otrzymywania materiałów i wykorzystywana do niej technika doboru próby generuje charakterystyczne rodzaje ubytków, AAPOR opracował kilka odpowiednich schematów klasyfikacyjnych, szczególnie opisujących możliwe „historie” każdej wylosowanej do próby jednostki. Niezależnie od technik, dla których zostały przygotowane, schematy te zawsze pozostają rozwinięciem czterech wymienionych wyżej kategorii.

CBOS posługuje się własną kategoryzacją do opisu przypadków niezrealizowania wywiadu i prezentacją wyników za pomocą miar AAPOR musi być

poprzedzona odpowiednim przekodowaniem danych. Tu pewnym ograniczeniem jest fakt, że AAPOR przygotował schemat klasyfikacyjny dla ubytków występujących w badaniach realizowanych techniką wywiadu *face to face* na próbach adresowych, ale nie imiennych. Ograniczenie to nie uniemożliwia jednak przeprowadzenia we własnym zakresie uporządkowania przypadków niezrealizowania wywiadu pod względem przynależności wylosowanych jednostek do badanej populacji. Brak gotowego schematu klasyfikacyjnego nie stanowi w tym przypadku zasadniczej przeszkody, tym bardziej, że – jak zauważa AAPOR w odniesieniu do prób imiennych (wykorzystywanych przy ankietach pocztowych lub internetowych) – o interpretacji przypadku ze względu na jego przynależność do badanej populacji decyduje w dużej mierze rodzaj i jakość operatu losowania [AAPOR 2012: 26]. Przykładowo, w sondażach realizowanych za pomocą osobistego wywiadu, niemożność skontaktowania się z kimkolwiek pod wskazanym adresem, przy niedokładnym operacie, będzie prowadziła do klasyfikacji „nie wiadomo, czy wylosowany respondent należy do badanej populacji” – nie można wykluczyć, że wylosowana osoba wcześniej zmarła lub przeprowadziła się w miejsce leżące poza obrębem badanej populacji. Natomiast ten sam przypadek niedostępności, ale zaistniały w próbie pobranej z aktualnego operatu pozwoli przyjąć zasadne założenie, że pod wskazanym adresem znajduje się osoba odpowiadająca kryteriom doboru. Przedstawiony przykład pokazuje, że konsekwencje w sposobie kodowania ubytków muszą być rozpatrywane zawsze pod kątem konkretnego operatu losowania.

W tabeli 1. prezentuję sposób, w jaki przyporządkowano wyróżniane przez CBOS rodzaje niedostępności (oddzielnie dla dwóch rodzajów stosowanych prób) do podstawowych kategorii schematu klasyfikacyjnego AAPOR. Mimo że możliwe jest operowanie w nim jedynie zgrubnymi kategoriami, mają one taki poziom szczegółowości, który umożliwia wyliczenie wartości skonstruowanych przez AAPOR współczynników realizacji.

Zmiana sposobu doboru próby z adresowej na imienną powoduje, że ta sama kategoria niedostępności może uzyskać inną klasyfikację ze względu na przynależność do badanej populacji. W próbie adresowej przypadki błędnych adresów (wskazany adres nie istnieje, pusty plac, pustostan, budynek niemieszkalny) są uznawane za jednostki spoza próby – nie należące do populacji. Inaczej rzecz się ma w przypadku prób imiennych, gdzie adres jest jedynie informacją pomocniczą, mającą umożliwić zlokalizowanie konkretnej osoby.

TABELA 1. Adaptacja schematów klasyfikacyjnych CBOS do głównych kategorii schematu AAPOR

|           | AAPOR   | CBOS od 1993 do sierpnia 2003<br>(próby adresowe)  | CBOS od września 2003<br>(próby imienne)   |
|-----------|---|--|--|
|           | UZYSKANO WYWIAD                                   |  |  |
| I         | W całości   | 17) wywiad zrealizowano  | wywiad zrealizowano  |
| P         | W części  |  |  |
|           | NALEŻY DO POPULACJI, BRAK WYWIADU                 |  |  |
| R         | Odmowy  | 3) odmowa wpuszczenia do mieszkania; 4) odmowa info. o osobach na stałe zamieszkujących; 10–16) wylosowany R. odmówił z powodu.... | 6) wylosowany R odmówił udziału w badaniu  |
| NC        | Nieemożność skontaktowania się                    | 2) pod wskazanym adresem nikt nie zastał; 6–8) wylosowany R był nieobecny....  | 5) wylosowany R nieobecny (nieuchwytny) przez cały okres realizacji badania  |
| O         | Inne  | 9) R niesprawny, ciężko chory  | 7) wylosowany R jest niesprawny w stopniu uniemożliwiającym przeprowadzenie wywiadu  |
|           | NIE WIADOMO CZY NALEŻY DO POPULACJI, BRAK WYWIADU |  |  |
| UH/<br>UO |   | 18) inne przyczyny niezrealizowania wywiadu; 0) braki danych   | 1) błędny adres; 2) pod wskazanym adresem nie można się z nikim skontaktować; 3) wylosowany R zmienił miejsce zamieszkania i nie można się z nim skontaktować; 4) wylosowany R nie żyje; 9) inne przyczyny niezrealizowania wywiadu; 0) braki danych |
|           | NIE NALEŻY DO POPULACJI                           |  |  |
|           |   | 1) wskazano niewłaściwy adres; 5) nie mieszka osoba odpowiadająca kryteriom doboru   | 8) wylosowany R nie spełnia kryteriów badania  |

Źródło: opracowanie własne.

W tym wypadku błąd w podanym adresie nie oznacza, że wylosowana osoba nie należy do populacji – jej status ze względu na tę przynależność jest w danym momencie niemożliwy do ustalenia. Różną klasyfikację będą mieć przypadki niemożności skontaktowania się z nikim w gospodarstwie domowym. W próbach adresowych mogły być traktowane jako należące do badanej populacji – do takiego ustalenia wystarczyło potwierdzenie, że lokal jest zamieszkały. Kryterium zamieszkania lokalu jest zbyt słabe dla prób imiennych, bo nie zapewnia, że mieszka tam osoba o konkretnym imieniu i nazwisku. Tu dodatkowo trzeba wziąć pod uwagę specyfikę bazy PESEL, w której adres jest przede wszystkim informacją o miejscu zameldowania, a nie faktycznego zamieszkania.

W przypadku prób imiennych komentarza wymaga zakwalifikowanie kategorii „wylosowany R odmówił udziału w badaniu” do grupy jednostek należących do populacji. Z punktu widzenia samej nazwy decyzja ta nie powinna budzić wątpliwości, jednakże uwzględniając realia sytuacji w terenie i fakt, że CBOS przewidział tylko jedną kategorię do rejestrowania odmów, można przypuszczać, że zawierają się tu także przypadki odmów udzielanych „na poziomie gospodarstwa domowego”, w tym przez osoby niewylosowane na respondenta. Abstrahując od nieadekwatności użytej nazwy, pytaniem jest czy odmowa komunikowana przez członka gospodarstwa domowego może przy okazji nieść informację na temat zamieszkiwania tam osoby wylosowanej na respondenta? Wydaje się, że takie założenie można przyjąć, biorąc pod uwagę, że rozmówca ma możliwość wyprowadzić ankietera z błędu w kwestii zamieszkiwania pod adresem osoby, której imię i nazwisko pada podczas aranżacji. To pośrednie rozumowanie prowadziło do wniosku, że jeśli rozmówca z tej możliwości nie korzysta, to oznacza, że poszukiwana przez ankietera osoba zamieszkuje wskazany lokal, a więc należy do badanej populacji.

Jeżeli chodzi o przypadki śmierci wylosowanego respondenta zdecydowano, że będą one kodowane do kategorii „nie wiadomo, czy należy do badanej populacji”. Standardy AAPOR mówią, że śmierć respondenta, o ile nastąpiła na przestrzeni okresu realizacji badania, należy traktować jako niedostępność jednostki należącej do populacji. Jednakże w CBOS nie zbiera się informacji o terminie tego zajścia. Jednocześnie wydaje się niemożliwe, aby operat PESEL i pobierane z niego próby charakteryzowały się tak wysokim stopniem aktualności, aby nie zawierały przypadków śmierci mających miejsce przed rozpoczęciem badania.

AAPOR opracował cztery podstawowe typy miar poziomu realizacji, a dla każdego typu możliwe jego warianty, oznaczone innymi symbolami: współczynnik odpowiedzi RR (w 6 wariantach); (w 3 wariantach); współczynnik odmów REF (w 3 wariantach) oraz współczynnik kooperacji COOP (w 4 wariantach)



[AAPOR 2011: 45–48]. Wyborem zastosowanych przeze mnie miar poziomu realizacji kierowała zasada, aby wynikiem, pochodzącym z dwóch rodzajów prób zapewnić jak największą porównywalność. Z tego względu dobrano takie rodzaje współczynników RR, CON i REF, które uwzględniają jednostki, o których nie wiadomo czy należą do badanej populacji – dzięki temu podstawy obliczeniowe przy obu rodzajach doboru stały się możliwie najbardziej podobne<sup>6</sup> i jednocześnie zbliżone do wielkości próby założonej. Rozwiązanie to ma jednak pewne wady w przypadku współczynnika kontaktów – miary, która została zaprojektowana do badania jednego z głównych typów niedostępności, jakim są nieobecności (jednostek należących do populacji). Przyjęty wariant współczynnika będzie dodatkowo odzwierciedlał zmiany zachodzące w zbiorze przypadków nie skontaktowania się z nikim z gospodarstwa domowego i wszystkich pozostałych, których status, ze względu na przynależność do badanej populacji, nie został ustalony. Z drugiej strony, przyjęte rozwiązanie jest jedynym, które umożliwia zbadanie wpływu doboru próby na współczynnik kontaktów, gdyż wymaga ono porównywalnych danych. Natomiast do badania zjawiska odmów najlepiej służy współczynnik kooperacji. Przydatność współczynnika odmów jest w tym względzie, wbrew nazwie, mniejsza, gdyż konstrukcja tej miary nie pozwala niezawodnie wnioskować o poziomie (nie)chęci do uczestniczenia w sondażach, będąc wypadkową zmian zachodzących w zbiorze odmów, nieobecności i innych przypadków niezrealizowania wywiadu. Z tego powodu większość analiz przeprowadzam z pominięciem tej miary. Dla współczynnika kooperacji wybrano wariant, który opiera się tylko na dwóch rodzajach reakcji – zgodach i odmowach uczestnictwa, i pozwala ustalić proporcję respondentów do tych jednostek, które obecne w domu, były zdolne do podjęcia tej roli. Przedstawiam wytypowane współczynniki oraz odpowiadające im wzory.

$$RR2 = \frac{I + P}{(I + P) + (R + NC + O) + (UH + UO)}$$

$$CON1 = \frac{(I + P) + R + O}{(I + P) + R + O + NC + (UH + UO)}$$

$$REF1 = \frac{R}{(I + P) + (R + NC + O) + (UH + UO)}$$

<sup>6</sup> Różnica między nimi związana z (nie)obecnością w podstawie oprocentowania przypadków błędnych adresów, których udział w próbie założonej jest jednak bardzo mały (średnia 0,5%) i względnie stały (min. 0,4%, max. 0,6%).

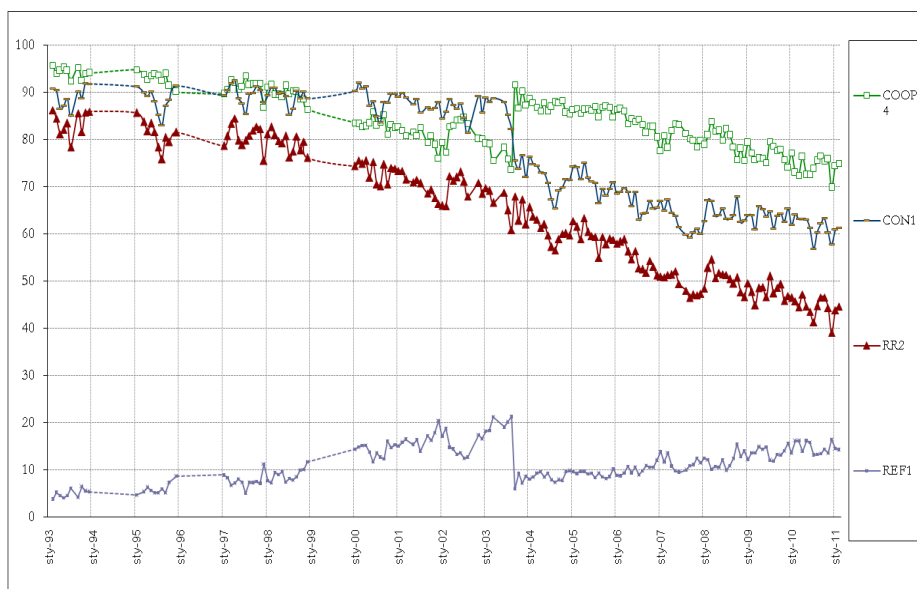
$$COOP4 = \frac{I + P}{(I + P) + R}$$

Z uwagi na to, że próby stosowane przez CBOS były próbami nieproporcjonalnymi, uzyskane dla kraju wartości współczynników realizacji kraju przeważono, wykorzystując do tego informacje o klasie miejscowości, w której wylosowano adres/jednostkę oraz publikowanej przez GUS strukturze ludności wg klas miejscowości dla poszczególnych roczników. Ten sposób ważenia z pewnością nie jest idealny, niemniej daje lepsze przybliżenie do wartości reprezentatywnych dla kraju niż opieranie się na danych pochodzących z prób, których strukturę wielokrotnie zmieniano. Potrzeba przeważenia danych jest związana także z podjętą próbą objaśnienia fluktuacji krzywej kooperacji, a czynniki wykorzystane w tej analizie mają status danych reprezentatywnych.

### TRENDY NIEDOSTĘPNOŚCI RESPONDENTÓW

Uszeregowane czasowo wartości, jakie przybierały w okresie 1993–2011 wybrane do prezentacji współczynniki realizacji przedstawia wykres 1. Pokazuje on monotoniczną tendencję spadkową w obrębie odsetka odpowiedzi (RR2) – średnie arytmetyczne w krańcowych latach 1993 i 2010 wynoszą odpowiednio 83,2 i 44,5%, co oznacza, że na przestrzeni 17 lat wartość tego współczynnika spadła o blisko połowę. Analiza krzywych CON1 i COOP4 pozwala uzyskać lepszy wgląd w mechanizm, który do tego doprowadził. Wyraźna zmiana w przebiegu tych krzywych następuje w 2003 r., a dokładnie od momentu zmiany sposobu doboru próby. Krzywa kontaktów, dotychczas przebiegająca średnio na poziomie 88%, w 2003 r. gwałtownie opadła o kilkanaście punktów, wykazując w późniejszym czasie mniejszą, ale systematyczną tendencję spadkową (średnia arytmetyczna w roku 2010 wyniosła 61,4%). Z kolei krzywa kooperacji, systematycznie opadająca w okresie stosowania prób adresowych, w 2003 r. rośnie o kilkanaście punktów, po czym, podobnie jak krzywa kontaktów, powoli zaczyna opadać (średnia arytmetyczna w latach 1993 i 2010, to odpowiednio 94,2% i 74,2%). Zmiana sposobu doboru próby wyznacza moment, w którym między krzywą kooperacji a krzywą kontaktów następuje „zamiana miejsc” – ta, która w parze uzyskiwała do tej pory wyższe wartości, obecnie przybiera mniejsze, i odwrotnie. Jak zatem widać, jednostajny spadek odsetka odpowiedzi RR2 wynika częściowo z podobnych, a częściowo z różnych, ale znoszących się tendencji w zbiorach odmów i nieobecności.

WYKRES 1. Wartości współczynników RR2, CON1, REF1 i COOP4 w latach 1993–2011 dla kraju

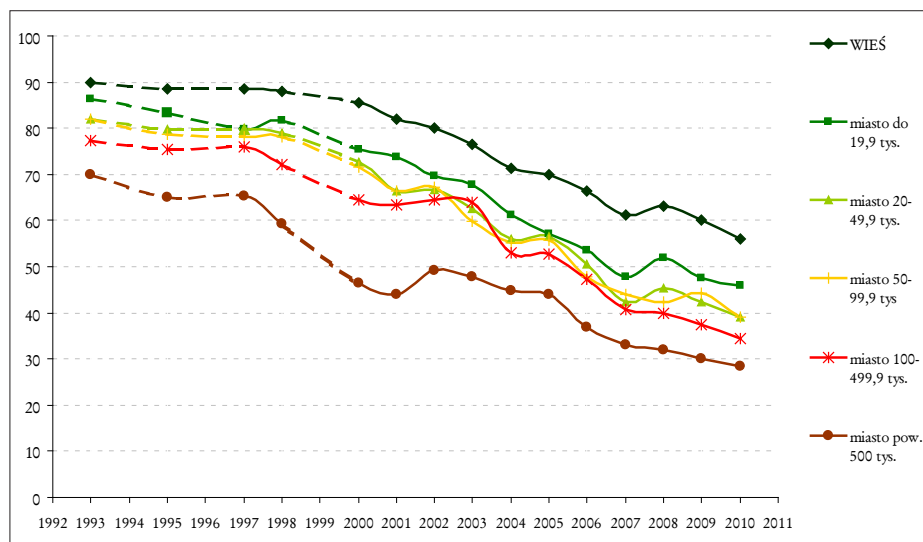


Źródło: opracowanie własne.

Prawidłowości te dokładniej przedstawia zamieszczona w aneksie tabela 3, która pokazuje związek między poszczególnymi współczynnikami w całym badanym okresie oraz w podokresach wyróżnionych ze względu na stosowany rodzaj doboru.

Wyniki uzyskane dla klas miejscowości przedstawiam graficznie na wykresach sporządzonych w oparciu o średnie roczne, a informację o ich wartościach zamieszczam w tabelach aneksowych (I i II). Wykres 2 obrazuje znaną prawidłowość, zgodnie z którą wartość odsetka odpowiedzi jest związana z wielkością miejscowości, w jakiej wylosowano jednostkę do badania. Zależność jest szczególnie widoczna w odniesieniu do skrajnych klas miejscowości, natomiast zmienna ta nie różnicuje tak silnie wartości RR2 w grupie małych i średnich miast.

WYKRES 2. Wartości średnich arytmetycznych współczynnika odpowiedzi RR2 w klasach miejscowości



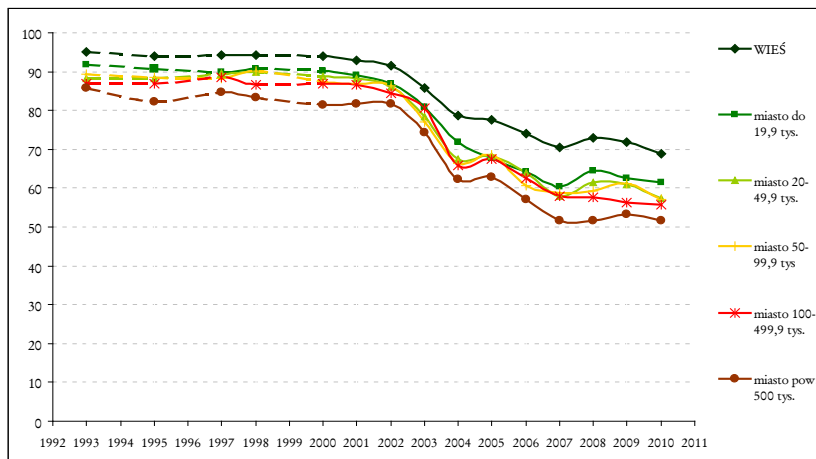
Źródło: opracowanie własne.

Krzywe kontaktów w poszczególnych klasach miejscowości (wykres 3) charakteryzuje podobny kształt, co oznaczałoby, że zjawisko niedostępności wynikające z nieobecności rozwija się wszędzie w zbliżony sposób, niemniej samo natężenie problemu jest zróżnicowane – występuje ujemna zależność między wartością CON1 a klasą miejscowości, tzn. najtrudniej jest zastać w domu mieszkańców dużych miast, a najłatwiej wsi.

W przypadku krzywych kooperacji (wykres 4), ujęcie przekrojowe unaocznia ten sam mechanizm wypadania jednostek z próby, który wystąpił w odniesieniu do krzywych kontaktów – szansa na pozyskanie wywiadu jest ujemnie związana z wielkością miejscowości. Z innych spostrzeżeń obserwuje się, że krzywe o wyższych wartościach mają bardziej stabilny przebieg – spadki i wzrosty o mniejszych zakresach. Podlegający największym zmianom jest kształt krzywej dla największych aglomeracji. Początek wzrostowej tendencji rozpoczął się już w 2002 r., a więc jeszcze przed zmianą techniki doboru próby, a prawidłowość ta zarysowała się w odniesieniu do wszystkich miast o większej liczbie ludności niż 19,9 tys. Największe odbicie nastąpiło jednak w wielkich miastach, choć wzrost ten nie był na tyle duży, aby COOP4 osiągnął poziom w tym czasie typowy dla miast o mniejszej liczbie ludności. To znaczące odwrócenie się niekorzystnej tendencji skoń-

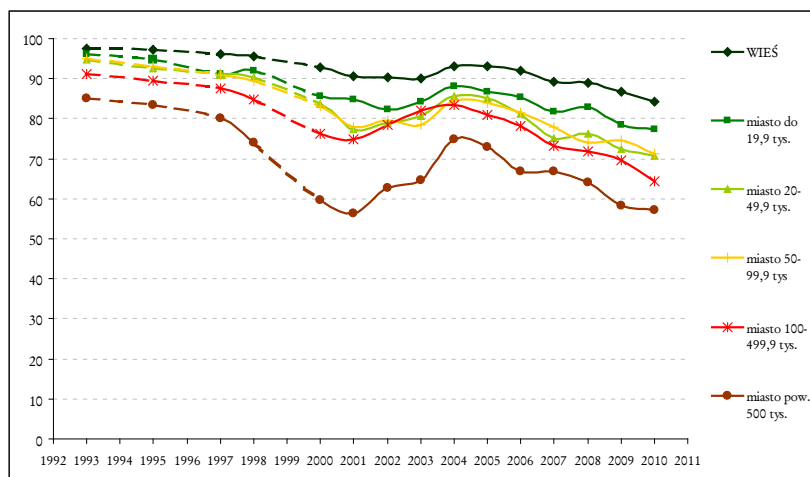
czyło się w 2005 r. – od tego momentu wszystkie krzywe przyjęły generalnie trend spadkowy, mimo krótkich okresów nieznacznego odbicia się albo wyhamowania.

WYKRES 3. Wartości średnich arytmetycznych współczynnika kontaktów CON1 w klasach miejscowości



Źródło: opracowanie własne.

WYKRES 4. Wartości średnich arytmetycznych współczynnika kontaktów COOP4 w klasach miejscowości



Źródło: opracowanie własne.

### CZYNNIKI WPLYWAJĄCE NA ZMIANY KRZYWEJ KONTAKTÓW W KRAJU

Mając w pamięci wzmiankowane już ograniczenia przyjętej miary CON1, podjęto ostrożną próbę udzielenia odpowiedzi na pytanie o czynniki wpływające na zmianę jej wartości w czasie. Wskazywane w literaturze przedmiotu czynniki dotyczące ujęcia przekrojowego to: istnienie fizycznych barier utrudniających nawiązanie kontaktu (np. zamknięte bramy); długość i częstość przebywania w domu; liczba wizyt ankietera i czas ich składania [Groves, Couper 1998: 80]. Ich przydatność do przeprowadzanej tu analizy wzdłuż czasowej jest ograniczona, co wynika z braku odpowiednich danych na temat utrudnień z wejściem do mieszkania oraz liczby wszystkich złożonych wizyt (ta ostatnia informacja zbierana jest dopiero od września 2003 r.). W literaturze spadki poziomu kontaktów tłumaczy się raczej długofalowymi zmianami w strukturze społeczno-demograficznej – wzrostem liczby jednoosobowych gospodarstw domowych, spadkiem liczby rodzin z dziećmi poniżej 6 roku życia, wzrostem udziału kobiet w rynku pracy itp. [Groves, Couper 1998: 90–93]. Oprócz zmian, które zaszły u potencjalnych respondentów, w bieżącej analizie przypuszczeniem najbardziej domagającym się sprawdzenia jest przede wszystkim wpływ sposobu doboru próby. Jeśli przebieg krzywej kontaktów miałyby wyjaśniać zmiana zachowań ankietowanych związana ze składaniem wizyt, to hipotezę tę można sprawdzić, uwzględniając zmienną mówiącą o średniej liczbie adresów przypadających na 1 ankietar<sup>7</sup> oraz zmienną mówiącą o wydłużeniu czasu realizacji badania o 2 dni, wiążącym się z wprowadzeniem CAPI [Bellos 2010].

W kształcie krzywej CON1 dla kraju zastanawia przede wszystkim przypadający na czas realizacji badań za pomocą prób imiennych, nagły, a potem postępujący wolniej wzrost liczby osób nieobecnych w trakcie wizyt(y) ankietera. Jak pokazują zamieszczone w tabeli 2 wyniki analizy korelacji, bardzo duży wpływ na obniżenie wartości CON1 miała zmiana doboru próby z adresowej na imienną. Obserwację tę można wyjaśnić „niewychwytywaniem” przez próby adresowe wszystkich przypadków dłuższych nieobecności, ponieważ przy tworzeniu operatu losowania ostatniego stopnia (listy osób zamieszkałych, zgłaszanej ankietarowi) obowiązywało kryterium faktycznego zamieszkiwania. Tym samym do próby nie wchodziły jednostki posiadające jedynie meldunek w lokalu, natomiast właśnie takie przypadki (np. dłuższych wyjazdów zagranicznych)

<sup>7</sup> Materiały z 1993 i 1995 roku odznaczały się niską jakością i zawierały braki danych dla zmiennej informującej o numerze ankietera, któremu przydzielono adres (maks. 188 braków w sondażu). Aby analizami objąć ten materiał, liczbę brakujących w zbiorze ankietarów ustalano dzieląc liczbę braków danych przez średnią przydzielonych adresów 1 ankietarowi (wyliczoną z komórek niepełnych).

„obciążają” próby imienne. Z kolei, spadki wartości CON1 w dalszym okresie można próbować tłumaczyć falą migracji zarobkowej do krajów UE, której tempo wzrostu było szczególnie wysokie od maja 2004 r., natomiast po 2007 r. nieco osłabło [Czapiński, Panek 2009: 123]. Z uwagi na to, że wyjazdy te miały przede wszystkim charakter czasowy i nie wiązały się ze zmianą obywatelstwa jednostki te pozostały w operacji losowania.

Druga z uwzględnionych zmiennych również bardzo silnie koreluje ze współczynnikiem CON1 – związek występuje w oczekiwanym kierunku i głosi, że spadek poziomu kontaktów następował wraz ze wzrostem liczby adresów przypadających na jednego ankietera (tabela 2). Warto dodać, że początek tendencji polegającej na coraz większym obciążeniu jednego ankietera pracą rozpoczął się około 2000 r. (średnia roczna 8,2 adresów). W roku 2008 przed wprowadzeniem CAPI liczba adresów wynosiła już średnio 12,6, a po wprowadzeniu aż 16,6, co związane było z początkowo ograniczoną liczbą laptopów [Bellos 2010]. W późniejszym czasie tę zmienną charakteryzowała nieznaczna tendencja spadkowa (15,4 w 2010 r.). Trzecia zmienna – wydłużenie czasu realizacji badania o 2 dni – ujawnia swój korzystny (choć słaby) wpływ dopiero w analizach wielowymiarowych przy kontrolowaniu pozostałych czynników.

Łącznie trzy włączone do analizy regresji zmienne wyjaśniają ponad 93% wariancji współczynnika kontaktów, co świadczy o trafnym doborze predyktorów.

TABELA 2. Wyniki analizy regresji przeprowadzonej metodą krokową

| Predyktor CON1 | Standaryzowany współczynnik regresji (Beta) | <i>p</i> | Współczynnik korelacji <i>r</i> | Korelacja cząstkowa | F       | Sumaryczna ocena modelu |                            |
|----------------|---|----------|---------------------------------|---------------------|---------|-------------------------|----------------------------|
|                |   |          |                                 |                     |         | <i>p</i>                | Skorygowany R <sup>2</sup> |
| $X_1$          | -0,703                                      | 0,000    | -0,947                          | -0,853              | 831,110 | 0,000                   | 0,936                      |
| $X_2$          | -0,414                                      | 0,000    | -0,813                          | -0,454              |         |                         |                            |
| $X_3$          | 0,116                                       | 0,019    | -0,559                          | 0,180               |         |                         |                            |

Legenda:  $X_1$  – rodzaj doboru próby (0 – adresowy, 1 – imienny);  $X_2$  – średnia liczba adresów przypadających na jednego ankietera;  $X_3$  – wydłużenie czasu realizacji (0 – nie, 1 – tak).

Źródło: opracowanie własne; N = 171.

### CZYNNIKI WPLYWAJĄCE NA ZMIANY KRZYWEJ KOOPERACJI W KRAJU

W przeciwieństwie do nieobecności, odmowy to o wiele bardziej skomplikowany fenomen. Poziom kooperacji jest funkcją cech sondażu, otoczenia społecznego, a także cech potencjalnego respondenta, ankietera i interakcji między nimi

w trakcie aranżacji wywiadu [Groves, Couper 1998: 30]. Jak już wspomniano, na przestrzeni badanego okresu dwie charakterystyki sondażu prowadzonego przez CBOS uległy zmianie. Są to sposób doboru próby oraz technika zbierania materiałów i wpływ tych zmian zostanie zbadany w pierwszej kolejności.

Przypuszczeniu, że rodzaj doboru nie jest obojętny dla poziomu kooperacji, trudno jest jednak nadać kierunek i wskazać, którego oddziaływanie byłoby bardziej niekorzystne. Przy próbach adresowych drażliwość związana z naruszaniem prywatności może budzić, potrzebne do sporządzenia operatu losowania 3 stopnia, pytanie ankietera o liczbę, płeć i wiek osób zamieszkałych w gospodarstwie domowym. Podobne reakcje może wywołać fakt dysponowania przez ankietera danymi tożsamościowymi przy próbach imiennych. Niewykluczone jednak, że drugi rodzaj doboru, w odniesieniu do publiczności nieprzejawiającej tego typu niepokojów, może mieć pozytywny efekt i w następstwie personalizacji próby wywoływać poczucie bycia wyróżnionym czy na swój sposób ważnym. Przedstawione w tabeli 3 wyniki analiz dwuzmiennych pokazują, że dobroczynny efekt próby imiennej ujawnia się tylko w pewnych podokresach, a szczególnie wtedy, gdy weźmie się pod uwagę kilka badań sprzed i po wprowadzeniu nowego rodzaju doboru. Efekt ten staje się statystycznie nieistotny po włączeniu danych odpowiadającym wszystkim próbom adresowym, co wynika z trwającej od 1993 r. spadkowej tendencji krzywej kooperacji. Z kolei uwzględnienie pełnego okresu daje rezultat mówiący o negatywnym oddziaływaniu tego czynnika na COOP4, co jest spowodowane powrotem krzywej do spadkowej tendencji w ostatnich latach. W celu ustalenia czy inne zmienne nie zakłóciły oddziaływania badanego tu czynnika, dalej zostanie przeprowadzona analiza wielowymiarowa. W tym miejscu można jedynie dodać, że zmiana doboru próby w wyróżnionych podokresach dała bardzo zbliżone efekty we wszystkich klasach miejscowości.

TABELA 3. Wartości korelacji  $r$  Pearsona między rodzajem doboru próby a współczynnikiem kooperacji COOP4 w wyszczególnionych okresach

|   |         |
|---|---------|
| Dla całości badanego okresu, $n = 171$                    | -0,45** |
| Do końca 2003 r. $n = 86$ (82 próby adresowe i 4 imienne) | 0,07    |
| W 2003 r. $n = 10$ (6 prób adresowych i 4 imienne)        | 0,95**  |
| Od 2003 r. $n = 95$ (6 prób adresowych i 89 imiennych)    | 0,22*   |

\*\* $p < 0,01$ ; \* $p < 0,05$ ; 0 – adresowa, 1 – imienna

Źródło: opracowanie własne.

Jeżeli chodzi o drugą z charakterystyk sondażu – technikę otrzymywania materiałów, która w badanym czasie uległa zmianie, to można wstępnie przypuszczać,



że jej wpływ, o ile w ogóle, nie będzie duży. Tak jak istotne różnice w odsetkach realizacji obserwuje się między wywiadem kwestionariuszowym, telefonicznym czy ankietą pocztową [Yu, Cooper 1983: 39; Hox, de Leeuw 1994: 330, 335, 339], tak wydaje się, że w przypadku wywiadu z kwestionariuszem w papierowej formie i jego wspomaganym komputerowo odpowiednikiem mogą one nie wystąpić. Oczekiwaniu temu towarzyszy przypuszczenie, że w procesie podejmowania decyzji odnośnie do uczestnictwa, element, jakim jest narzędzie pracy ankietera, nie ma zasadniczego znaczenia dla potencjalnego respondenta. Spodziewany niewielki efekt oddziaływania tego czynnika może też być konsekwencją występowania u ludzi reakcji o przeciwnych kierunkach, które będą znosić się w zbiorze. Jak bowiem wiadomo na podstawie wyników badań w innych krajach, obecność komputera budzi ambiwalentne reakcje w kwestii poczucia prywatności. Z drugiej strony, wskazuje się na pozytywne wartościowanie postępu technicznego, które powoduje, że użycie komputera w wywiadzie jawi się pozytywnie w oczach potencjalnego respondenta i przekłada się na obraz badania, ankietera i siebie samego [Sztabiński P. B. 2001]. Należy jeszcze dodać, że w bieżącej analizie wpływ CAPI nie będzie mógł być uchwycony w „czystej” postaci, a to z tego względu, że wraz ze zmianą techniki w CBOS miały miejsce dwie niewielkie modyfikacje: podwyższenie stawki za zrealizowany wywiad (o średnio 4 zł, co stanowi ok. 19% poprzedniego wynagrodzenia) oraz zmniejszenie czasu trwania wywiadu ze średnio 52 do 42 minut (efekt zmniejszenia objętości kwestionariusza) [Bellos 2010].

W analizie dwuzmiennej okazuje się jednak, że CAPI ma negatywny wpływ na tę zmienną niezależnie od okresu, w jakim rozpatruje się ten związek. Zależność jest szczególnie silna, jeśli weźmie się pod uwagę kilka badań sprzed i po wprowadzeniu nowej techniki otrzymywania materiałów, ale traci na sile w dłuższym okresie – w każdym jednak przypadku pozostaje na wyższym poziomie niż się pierwotnie spodziewano (tabela 4).

TABELA 4. Wartości korelacji  $r$  Pearsona między techniką otrzymywania materiałów a współczynnikiem kooperacji COOP4 w wyszczególnionych okresach

|   |         |
|---|---------|
| Dla całości badanego okresu $n = 171$                     | -0,63** |
| Do końca 2008 r. $n = 145$ (141 prób PAPI i 4 próby CAPI) | -0,28** |
| W 2008 r. $n = 12$ (8 prób PAPI i 4 próby CAPI)           | -0,83** |
| Od 2008 r. $n = 38$ (8 prób PAPI i 30 prób CAPI)          | -0,74** |

\*\* $p < 0,01$ ; (0 – PAPI, 1 – CAPI)

Źródło: opracowanie własne.

Interpretację tego wyniku utrudnia fakt, że reakcja na CAPI była podobna we wszystkich klasach miejscowości, w efekcie trudno wiązać go z potrzebą ochrony prywatności (z innych badań wiadomo, że wielkość miejscowości zamieszkania znacząco wpływa na tę cechę) czy z wyobrażeniem o doniosłości i wiarygodności przedsięwzięcia, jakie wywołuje ankietę z laptopem (można się spodziewać, że wielkość miejscowości zamieszkania również je różnicuje).

Skoro przyczyna ma „uniwersalny” charakter, można się zastanowić, czy nie tkwi ona we wzorze postępowania ankietatorów i wiąże się z podejrzeniem, że na skutek zwiększenia wynagrodzenia mogli oni pracować z nieco mniejszą determinacją w pozyskiwaniu respondentów (wkładając mniej wysiłków mogli uzyskać ten sam zarobek). Przypuszczenie to nie wydaje się jednak prawdziwe w świetle wyników, które mówią o tym, że średnia bezwzględnych odsetków zwrotów, jakie osiągnęli poszczególni ankietery pracujący zarówno w ostatnim badaniu PAPI jak i pierwszym CAPI, nie zmieniła się istotnie wraz z przejściem na nową technikę (odpowiednio 48,3% i 49,0%). Na zmiany we wzorach zachowania nie wskazują także wartości odchylenia standardowego (odpowiednio 22,5% i 19,0%). Na przestrzeni sondaży realizowanych w 2003 r. nie uległa też zmianom średnia liczba wizyt kończących się wywiadem oraz średnia liczba wizyt kończących się odmową<sup>8</sup>. Przedstawione tu rezultaty badań nad wpływem zmiany stosowanej techniki otrzymywania materiałów skłaniają do podjęcia dalszych, tym razem wielowymiarowych analiz, które pozwolą sprawdzić, czy uzyskane wyniki utrzymają się, jeśli oddziaływanie innych zmiennych będzie kontrolowane.

Chociaż w literaturze zagranicznej trudno zetknąć się z hipotezą mówiącą, iż za wzrostem odmów stoi pogarszająca się jakość pracy ankietatorów, przypuszczenia takiego być może nie należy zbyt szybko odrzucać w odniesieniu do polskich realiów. Należy tu chociażby wziąć pod uwagę to, że instytucjom badawczym nie udaje się zapewnić takich warunków pracy, aby stworzyć na miarę swoich potrzeb stały zespół profesjonalnych ankietatorów, świadczących rzetelną pracę (relatywnie niskie zarobki, dorywczy charakter pracy). Wyobrażenie na temat zapotrzebowania na pracowników i ich rotacji dają statystyki dotyczące ankietatorów CBOS w badaniach APIW w okresie wrzesień 2003 – luty 2011. W tym czasie łącznie pracowało ich 1511. Ankietery, których kontakt z pracą był praktycznie jednorazowy (wzięli maksymalnie 12 adresów) stanowią aż 25% całej grupy. Połowa otrzymała do realizacji maksymalnie 32 wywiady. W trzeciej ćwiartce znajdują się osoby, które podjęły się realizacji do 108 wywiadów, a ostatnią stanowią stali

<sup>8</sup> Porównanie nie uwzględnia pomiarów z września 2008 (pierwsze zastosowanie CAPI), gdyż nie zbierano wtedy informacji na temat liczby złożonych wizyt.

współpracownicy (maksymalna liczba przydzielonych adresów to 2048). Niestety, nie istnieje zbyt wiele danych, które mogłyby zostać wykorzystane jako wskaźniki jakości pracy ankietera do przeprowadzenia analizy dla całości badanego okresu. Jak już wspomniano, informacja o ilości złożonych przez ankietera wizyt pod adresem jest zbierana dopiero od 2003 r. W tej sytuacji do wykorzystania pozostaje zmienna, która mówi o obciążeniu pracą ankietera, a mierzona jest liczbą przydzielonych adresów (lub jednostek), w których należy zrealizować wywiad. Jak pokazano we wcześniejszym fragmencie od poziomu tego obciążenia zależy z iloma jednostkami uda się skontaktować ankieterowi (lub inaczej – jaki będzie udział nieobecności w zbiorze), a teraz zostanie sprawdzony wpływ tej zmiennej na odsetek pozytywnych reakcji na prośbę o udział w badaniu wśród jednostek skontaktowanych i zdolnych do podjęcia roli respondenta (lub inaczej – na udział odmów w zbiorze). Za analizami tymi w obu przypadkach stoi zdroworozsądkowe przypuszczenie, że wraz ze zwiększeniem ilości wykonywanej pracy obniża się jej jakość. Wyniki analiz potwierdzają ten oczekiwany rezultat – wartość korelacji  $r$  Pearsona między średnią liczbą przydzielanych adresów w sondażu, a współczynnikiem COOP4 wynosi  $-0,72$  ( $p < 0,01$ ) (tabela 5).

Inne stawiane w literaturze hipotezy, mające objaśniać trendy współczynnika kooperacji, podkreślają zmiany, jakie zachodzą po stronie potencjalnych respondentów. Istnieje grupa hipotez, która zmiany w zakresie kooperacji wiąże przede wszystkim ze zmieniającą się strukturą demograficzną społeczeństw. Ustalenia GUS, jeśli chodzi o liczbę mieszkańców w poszczególnych klasach miejscowości, nie wskazują na żadne istotne przesunięcia [GUS 2011], które miałyby tłumaczyć tak dużą zmianę wartości współczynnika kooperacji. Inne przypuszczenie koncentruje się na problemie wzrostu lęku przed przestępczością, czyniąc je głównym powodem zmniejszania się chęci do udziału w badaniach. Brak odpowiednich danych uniemożliwia systematyczne ustosunkowanie się do tej hipotezy, istnieją jednak pewne przesłanki, które czynią ją mało prawdopodobną [Grzeszkiewicz-Radulska 2009: 182–185]. Z tego samego powodu trudna do przeprowadzenia jest ścisła weryfikacja hipotezy o wzroście potrzeby ochrony prywatności – dane o wysokim poziomie agregacji wskazują na dogodne warunki do rozwoju tego zjawiska w Polsce, a jakościowe świadectwa zebrane w prasie w związku z przeprowadzaniem spisów powszechnych pozwalałyby na ostrożne przychylenie się do tej hipotezy [Grzeszkiewicz-Radulska 2009: 201–219]. Jeszcze mniejsze możliwości sprawdzenia dotyczą hipotez przesylenia sondażami oraz narastania problemów związanych z podszywaniem się pod działalność sondażową przez różne podmioty prowadzące sprzedaż, telemarketing czy zbiórki pieniędzy [Grzeszkiewicz-Radulska 2009: 219–222]. Inne, możliwe do przetestowania

przypuszczenie głosi, że wzrost odmów w badaniach jest przejawem szerszego zjawiska, jakim jest apatia polityczna, słabnięcie norm w zakresie powinności obywatelskich, erozja idei społecznej kooperacji i działania na rzecz wspólnego dobra. Hipoteza ta oparta jest na przekonaniu, że akt partycypacji sondażowej (przede wszystkim w badaniach opinii publicznej lub realizowanych na zlecenie władz), podobnie jak akt partycypacji wyborczej, jest rodzajem aktywności obywatelskiej. Przyjęcie tego założenia uzasadnia poszukiwanie czynników wpływających na decyzje odmowne wśród korelatów absencji wyborczej. Przełożenie tej hipotezy z ujęcia przekrojowego na wzdłuż czasowe niesie jednak problemy, gdyż wiąże się z dostarczeniem dowodu na drodze ekologicznych analiz. Polegają one na odtworzeniu kontekstu makrospołecznego poprzez zaprezentowanie zmieniających się w czasie wartości zmiennych, które objaśniają absencję wyborczą, a następnie sprawdzeniu, w jakiej mierze kontekst ten wpływa na zmienność krzywej kooperacji. Analiza ma przynieść odpowiedź na pytanie, czy obserwuje się prawidłowości typowe dla głosowania wyborczego, jeśli chodzi o kierunek zależności [Kojetin, Tucker, Cashman 1994; Kojetin, Tucker 1999; [Grzeszkiewicz-Radulska 2009: 186–200].

Niestety nie udało się pozyskać informacji o wartościach zmiennych niezależnych odnoszących się do sfery świadomościowej (oceny, opinie) z innego źródła niż CBOS. Z uwagi jednak na fakt przeważenia tych danych, przyjmuję założenie o ich reprezentatywności dla badanej populacji. Ograniczeniem analiz jest też brak danych dla wielu zmiennych, o których z literatury przedmiotu wiadomo, że są korelatami absencji wyborczej.

Podobne analizy prowadzone dla okresu 1993–2003 dały rezultat mówiący o znaczącym, dodatnim wpływie nastrojów społecznych i opinii na temat naczelnych instytucji władz na odsetek kooperacji. Jednocześnie, kontrolując oddziaływanie pozostałych czynników, zaobserwowano, że odsetek ten spada wraz z polepszaniem się poziomu życia respondenta i jego rodziny [Grzeszkiewicz-Radulska 2009: 195–198].

Z kolei, wyniki analizy obejmującej pełny okres (1993–2011) pokazują, po pierwsze, mniejszą rolę czynników, jakimi są oceny sytuacji w kraju i działalności władz w objaśnieniu zmienności krzywej kooperacji. Związek między COOP4 a zmiennymi mierzącymi opinie i prognozy w odniesieniu do ogólnej oraz politycznej sytuacji w kraju charakteryzują niskie, nie zawsze istotne statystycznie, ale o oczekiwany dodatnim kierunku, wartości współczynnika korelacji<sup>9</sup>.

<sup>9</sup> Z uwagi na mniejszą liczbę pomiarów wyeliminowana z dalszych analiz zmienna „pozytywna ocena działalności Sejmu” koreluje na poziomie  $r = 0,32$  ( $n = 143$ ;  $p < 0,01$ ). Z kolei, ani stosunek

Z powyższej prawidłowości wyłamuje się jednak większość zmiennych odnoszących się do wymiaru gospodarczego w skali kraju oraz poziomu życia respondenta i jego rodziny – dobra ocena bieżącej sytuacji działa zniechęcająco do uczestnictwa, podobnie jak dobra sytuacja na rynku pracy mierzona stopą bezrobocia. Zwraca przy tym uwagę, że liczba osób pozytywnie oceniających poziom swojego życia (i warunki materialne gospodarstwa domowego<sup>10</sup>) jest zmienną najsilniej związaną ze współczynnikiem kooperacji (w grupie korelatorów absencji wyborczej, które w tabeli 5. zaznaczono ciemniejszym kolorem). Na tym tle odstają jednak zmienne odnoszące się do prognoz, a szczególnie tej związanej z sytuacją gospodarczą w skali kraju – optymizm w tym zakresie, podobnie jak w zakresie ogólnej sytuacji i politycznej w Polsce, przekłada się na wyższy poziom kooperacji, choć nie jest to związek silny.

Do analizy regresji użyto czynników, dla których liczba pomiarów była największa, tj. równa pomiarom zmiennej zależnej ( $n = 171$ ) oraz między którymi nie zachodził związek wyższy bądź równy 0,9. Włączono do niej wszystkie zmienne, których wpływ dotąd rozpatrywano. Wyniki zawiera tabela 6.

Do równania regresji weszły dwie zmienne pozostające pod kontrolą badacza: rodzaj próby oraz poziom obłożenia ankietowanych pracą. Przy kontroli pozostałych uwzględnionych czynników okazało się, że przejście z prób adresowych na imienne było decyzją korzystną. W przestrzeni wielowymiarowej stracił na sile, choć nie zmienił kierunku, związek między COOP4 a średnią liczbą przydzielanych adresów jednemu ankietowemu. Wyniki ujawniają spodziewany brak wpływu zmiany narzędzia (kwestionariusz w papierowej formie vs laptop) na poziom kooperacji.

Bardzo silnym predyktorem obniżającym prawdopodobieństwo partycypacji pozostaje liczba osób, które dobrze oceniają materialny poziom życia w swojej rodzinie. Kontrola pozostałych czynników spowodowała jednak, że wpływ stopy bezrobocia stał się niezgodny z tą prawidłowością – im jest ona wyższa, tym mniej chętnych do udzielania wywiadów<sup>11</sup>. Partycypacji nie sprzyja przy tym negatywna ocena sytuacji w kraju.

---

do rządu, ani ocena jego polityki gospodarczej nie są związane istotnie statystycznie z wartościami COOP4 ( $n = 167$ ).

<sup>10</sup> Między oceną poziomu życia respondenta i jego rodziny a oceną warunków materialnych w jego gospodarstwie domowym zachodzi korelacja bliska jedności, co świadczyłoby o tym, że zmienne te mierzą praktycznie tę samą cechę. Do analiz została wybrana ocena poziomu życia – zmienna o pełnej liczbie pomiarów.

<sup>11</sup> Być może obserwację tę należy wiązać z rozwarstwieniem społeczeństwa polskiego lub niską trafnością stopy bezrobocia, będącej wskaźnikiem ekonomicznej sytuacji jednostek, o czym

TABELA 5. Macierz korelacji dla zmiennych użytych w analizie regresji

|                 |                |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |                 |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | X <sub>1</sub> | X <sub>2</sub> | X <sub>3</sub> | X <sub>4</sub> | X <sub>5</sub> | X <sub>6</sub> | X <sub>7</sub> | X <sub>8</sub> | X <sub>9</sub> | X <sub>10</sub> | X <sub>11</sub> | X <sub>12</sub> | X <sub>13</sub> | X <sub>14</sub> | X <sub>15</sub> |
| X <sub>1</sub>  | 1              | -0,45**        | -0,63**        | -0,15*         | -0,13          | -0,13          | 0,25**         | -0,16*         | -0,27**        | 0,19*           | -0,67**         | 0,00            | 0,27**          | 0,12            | -0,72**         |
| X <sub>2</sub>  |                | 1              | 0,44**         | 0,02           | -0,18*         | 0,40**         | -0,04          | 0,06           | 0,49**         | 0,22**          | 0,72**          | 0,24**          | -0,55**         | -0,05           | 0,71**          |
| X <sub>3</sub>  |                |                | 1              | -0,09          | 0,04           | -0,06          | -0,11          | 0,17*          | 0,35**         | -0,01           | 0,72**          | 0,23**          | -0,21**         | -0,44**         | 0,88**          |
| X <sub>4</sub>  |                |                |                | 1              | 0,75**         | 0,79**         | -0,69**        | 0,67**         | -0,64**        | -0,79**         | -0,33**         | -0,83**         | 0,50**          | 0,60**          | -0,04           |
| X <sub>5</sub>  |                |                |                |                | 1              | 0,43**         | -0,68**        | 0,77**         | -0,61**        | -0,80**         | -0,30**         | -0,79**         | 0,78**          | 0,36**          | -0,06           |
| X <sub>6</sub>  |                |                |                |                |                | 1              | -0,51**        | 0,51**         | -0,35**        | -0,43**         | -0,12           | -0,57**         | 0,13            | 0,57**          | 0,13            |
| X <sub>7</sub>  |                |                |                |                |                |                | 1              | -0,65**        | 0,36**         | 0,77**          | 0,06            | 0,70**          | -0,39**         | -0,27**         | -0,15*          |
| X <sub>8</sub>  |                |                |                |                |                |                |                | 1              | -0,37**        | -0,68**         | -0,11           | -0,56**         | 0,48**          | 0,21**          | 0,12            |
| X <sub>9</sub>  |                |                |                |                |                |                |                |                | 1              | 0,65**          | 0,78**          | 0,76**          | -0,68**         | -0,74**         | 0,49**          |
| X <sub>10</sub> |                |                |                |                |                |                |                |                |                | 1               | 0,31**          | 0,83**          | -0,60**         | -0,40**         | 0,06            |
| X <sub>11</sub> |                |                |                |                |                |                |                |                |                |                 | 1               | 0,48**          | -0,61**         | -0,54**         | 0,85**          |
| X <sub>12</sub> |                |                |                |                |                |                |                |                |                |                 |                 | 1               | -0,68**         | -0,62**         | 0,27**          |
| X <sub>13</sub> |                |                |                |                |                |                |                |                |                |                 |                 |                 | 1               | 0,24**          | -0,40**         |
| X <sub>14</sub> |                |                |                |                |                |                |                |                |                |                 |                 |                 |                 | 1               | -0,37**         |
| X <sub>15</sub> |                |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 | 1               |

LEGENDA: X<sub>1</sub> – współczynnik kooperacji COOP4; X<sub>2</sub> – rodzaj doboru próby (0-adresowy, 1-imienny); X<sub>3</sub> – technika otrzymywania materiałów (0-PAPI, 1-CAPI); X<sub>4</sub> – zła ocena sytuacji w kraju; X<sub>5</sub> – prognoza pogorszenia sytuacji w kraju; X<sub>6</sub> – zła ocena sytuacji politycznej w kraju; X<sub>7</sub> – prognoza poprawy sytuacji politycznej w kraju; X<sub>8</sub> – prognoza pogorszenia sytuacji politycznej w kraju; X<sub>9</sub> – dobra ocena sytuacji gospodarczej w kraju; X<sub>10</sub> – prognoza poprawy sytuacji gospodarczej w kraju; X<sub>11</sub> – dobra ocena poziomu życia respondenta i jego rodziny; X<sub>12</sub> – prognoza poprawy poziomu życia respondenta i jego rodziny; X<sub>13</sub> – prognoza pogorszenia poziomu życia respondenta i jego rodziny; X<sub>14</sub> – stopa bezrobocia; X<sub>15</sub> – średnia liczba adresów przypadających na jednego ankietera.

Źródło: Opracowanie własne; N = 171.

TABELA 6. Wyniki analizy regresji przeprowadzonej metodą krokową

| Predyktor COOP4 | Standaryzowany współczynnik regresji (Beta) | p     | Współczynnik korelacji r | Korelacja cząstkowa | F      | Sumaryczna ocena modelu |                            |
|-----------------|---|-------|--------------------------|---------------------|--------|-------------------------|----------------------------|
|                 |   |       |                          |                     |        | p                       | Skorygowany R <sup>2</sup> |
| X <sub>15</sub> | -0,38                                       | 0,000 | -0,72                    | -0,32               | 94,381 | 0,000                   | 0,733                      |
| X <sub>11</sub> | -1,10                                       | 0,000 | -0,67                    | -0,63               |        |                         |                            |
| X <sub>14</sub> | -0,42                                       | 0,000 | 0,12                     | -0,45               |        |                         |                            |
| X <sub>2</sub>  | 0,60  | 0,000 | -0,45                    | 0,54                |        |                         |                            |
| X <sub>4</sub>  | -0,30                                       | 0,000 | -0,15                    | -0,39               |        |                         |                            |

LEGENDA: jak w tabeli 5

Źródło: opracowanie własne; N = 17.

świadczy umiarkowana zależność między wielkością stopy bezrobocia a liczbą osób dobrze oceniających poziom swojego życia (r = -0,54; p < 0,01).

Porównując rezultaty osiągnięte w pierwszym okresie (1993–2003) do tych z pełnego okresu (1993–2011) w objaśnieniu fluktuacji krzywej COOP4, należy stwierdzić osłabienie roli czynników, jakimi są nastroje społeczne oraz wzmocnienie wcześniej zarysowanej tendencji, zgodnie z którą dobrostan ekonomiczny jednostek zmniejsza chęć partycypacji w sondażach. Ogólna interpretacja tych wyników będzie jednak podobna – czynniki objaśniające niechęć do kooperacji okazują się stanowić zasadniczy rys kondycji współczesnej polskiej demokracji. Zgodnie z ustaleniami Diagnostyki 2011:

Nie wyruszyliśmy jeszcze w drogę ku społeczeństwu obywatelskiemu. Ale rozwijamy się, i to w całkiem niezłym tempie, [...], tyle że w ostatnich latach dużo szybciej indywidualnie niż zespołowo. [...] Polacy opanowali niezłe sztukę gry z państwem i w związku z tym widzą coraz słabszy związek między tym, co robi państwo (władze), a tym, jak wygląda ich własne życie. Zaradność Polaków umożliwia im poprawę własnego bytu bez oglądania się na innych i niezależnie od stanu wspólnoty. [...] Mimo systematycznej poprawy indywidualnego bytu (niemal trzykrotny spadek ubogich gospodarstw od 1992 r.) pozostajemy niezadowoleni z sytuacji w kraju (niemal taki sam poziom jak w 1997 r., konsekwentnie od początku transformacji jest to najniższy wskaźnik satysfakcji w zbiorze około dwudziestu różnych aspektów życia – [...]) [Czapiński, Panek 2011: 353].

Ostatecznie, nie wszystkie zmienne objaśniające absencję wyborczą korelowały w kierunku typowym dla zależności ustalanych w analizach przekrojowych, ale jak widać odkryte prawidłowości nie kłócą się z doniesieniami o stanie polskiej demokracji i mechanizmach jej rozwoju. Co więcej, kierunki odkrytych zależności pozostają w zgodzie z wynikami analizy Kojetina i Tuckera, którzy przeprowadzili ją w oparciu o dane dla *Current Population Survey* za lata 1960–1988 i ustalili, że wzrost odmów następuje wraz z pogorszeniem nastrojów społecznych oraz poprawą sytuacji ekonomicznej [Kojetin, Tucker 1999: 177, 180].

Rezultaty przeprowadzonych tu badań analizy pokazują wyraźnie, że zjawisko niedostępności respondentów rozwija się w bardzo szybkim tempie, stanowiąc coraz większe wyzwanie dla empiryków realizujących badania sondażowe, w szczególności na próbach losowych. Problem ten jest bezpośrednio mniej widoczny w przypadku badań wykorzystujących inne doборы, niemniej i tam, nawet przy zrealizowaniu badania na próbie o założonych kryteriach, nie uzyskuje się gwarancji, że pod względem pozostałych cech próbka nie jest specyficzną, dającą wypaczone rezultaty, grupą osób. Narastanie trudności zdaje się wynikać w dużej części ze zmian, jakie zaszły po stronie potencjalnych respondentów. Chociaż głoszące to hipotezy trudno poddać rygorystycznym testom, tę argumentację wspierają liczne świadectwa zebrane w badaniach przekrojowych, a także trendy niedostępności obserwowane w innych cyklicznie prowadzonych sondażach. W Polsce jest praktycznie tylko jedna taka inicjatywa badawcza,

której wyniki można, aczkolwiek z wielką ostrożnością, zestawiać z sondażami CBOS – to badania prowadzone przez Polski Generalny Sondaż Społeczny przy Uniwersytecie Warszawskim. Poziom realizacji próby w latach 1992, 1993, 1994, 1995, 1997, 1999, 2002, 2005, 2008 wyniósł tu odpowiednio 82,4; 82,5; 80,5; 80,2; 75,1; 67,0; 61,7; 60,6; 51,8 [%]<sup>12</sup>. Jak widać, tendencja jest ta sama, a wielkości porównywalne.

W zakończeniu warto wyeksponować te rezultaty, w świetle których istnieje dla badaczy pewne pole do działania i szansa na walkę ze spadającymi odsetkami realizacji. Przypomnę, że szukając czynników objaśniających spadki współczynnika kontaktów oraz kooperacji ustalono, że ich silnym predyktorem okazała się średnia liczba adresów przydzielanych 1 ankierowi – im była ona większa tym ankierzy uzyskiwali gorsze noty zarówno w odnajdywaniu potencjalnych respondentów, jak i przekonywaniu ich do partycypacji. Waga tego ustalenia zwiększy się, jeśli uwzględnimy wynik wskazujący na istnienie zależności między wartościami współczynnika kontaktów i kooperacji (tabela aneksowa 3), co świadczy o tym, że osoby łatwe i trudne do skontaktowania charakteryzuje różna chęć kooperacji. Dodatni kierunek związku wskazuje, że jest ona większa w tej drugiej grupie. Na koniec proponowałabym spojrzeć na omawianą tu zmienną pod kątem warunków pracy stwarzanych ankierom przez badaczy, czy jeszcze szerzej – wskaźnik problemów ze znalezieniem na rynku rzetelnych i profesjonalnych współpracowników. Być może agencje badawcze powinny pomyśleć o możliwości reorganizacji i zmiany polityki zatrudniania ankierów. Trudno nie zgodzić się, że ta dotychczasowa sprzyja myśleniu o pracy ankierskiej jak o dorywczej, często w kategoriach zła koniecznego, lokującej się na pozycjach prestiżowo niewiele wyższych od akwizycyjnej. Pozyskać oddanych i sprawnych ankierów poprzez podniesienie atrakcyjności ich pracy i profesjonalne do niej przygotowanie – jak pokazują spadające na świecie odsetki realizacji, niewiele więcej jesteśmy w stanie zrobić.

---

<sup>12</sup> [www.pgss.iss.uw.edu.pl](http://www.pgss.iss.uw.edu.pl) [dostęp wrzesień 2011].



**BIBLIOGRAFIA**

- Bellos A. [2010], *Porównanie warunków i wskaźników realizacyjnych w badaniach realizowanych metodą tradycyjną (PAPI) i przy użyciu laptopów (CAPI)*, referat wygłoszony na IV Zjeździe Socjologicznym w Krakowie.
- Czapiński J., Panek T. [2009], *Diagnoza społeczna 2009*, [www.diagnoza.com](http://www.diagnoza.com) [dostęp sierpień 2011].
- Czapiński J., Panek T. (red.) [2011], *Diagnoza Społeczna 2011*, [www.diagnoza.com](http://www.diagnoza.com) [dostęp wrzesień 2011].
- Dąbala W. [2007], *Próby CBOS oraz szacowanie na ich podstawie parametrów populacji w badaniach sondażowych po roku 2000*, „Przegląd Socjologiczny”, t. LVI/1.
- Główny Urząd Statystyczny, *Ludność i struktura w przekroju terytorialnym*, [www.stat.gov.pl/gus](http://www.stat.gov.pl/gus) [dostęp lipiec 2011].
- Groves R., Couper M. [1998], *Nonresponse in household interview surveys*, New York: John Wiley & Sons.
- Grzeszkiewicz-Radulska K. [2009], *Respondenci niedostępni w badaniach sondażowych*, [w:] *Analizy i próby technik badawczych w socjologii*, t. XII, Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Hox J.J., Leeuw E.D. de [1994], *A comparison of nonresponse in mail, telephone, and face-to-face surveys*, “Quality and Quantity”, vol. 28, no. 4.
- Kojetin B., Tucker C. [1999], *Exploring the relation of economic and political conditions with refusal rates to government survey*, “Journal of Official Statistics”, vol. 15, no. 2.
- Kojetin B., Tucker C., Cashman E. [1994], *Response to a government survey as political participation: the relation of economic and political conditions to refusal rates in the Current Population Survey*, Proceedings of the American Statistical Association, Survey Methods Research Section, American Statistical Association.
- Lynn P., Beerten P., Laiho J.M. [2001], *Recommended standard final outcome categories and standard definitions of response rate for social surveys*, Working papers of the Institute for Social and Economic Research, paper 2001-23, Colchester: University of Essex.
- Polski Generalny Sondaż Społeczny, [www.pgss.iss.uw.edu.pl](http://www.pgss.iss.uw.edu.pl) [dostęp wrzesień 2011].
- Slater M., Christensen H. [2002], *Applying AAPOR'S final disposition codes and outcome rates to the 2000 Utah Colleges' exit poll*, American Association for Public Research, Strengthening Our Community – Section on Survey Research Methods.
- Sztabiński P.B. [2001], *Wywiad osobisty ze wspomaganie komputerowym (CAPI): czy rzeczywiście idealna technika?*, „ASK”, vol. 10.
- The American Association for Public Opinion Research [2011], *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 7th edition AAPOR.
- Yu J., Cooper H., [1983], *A quantitative review of research design effects on response rates to questionnaires*, “Journal of Marketing Research”, vol. 20.

**ANEKS**

TABELA I. Wartości statystyk opisowych dla poszczególnych współczynników realizacji na poziomie kraju

| Rok                                | $\bar{y}$ | $\sigma$ | Roczne tempo wzrostu(+) / spadku(-) | Średnie roczne tempo wzrostu(+) / spadku(-)* |
|------------------------------------|-----------|----------|-------------------------------------|--|
| 1                                  | 2         | 3        | 4                                   | 5  |
| <b>WSPÓLCZYNNIK ODPOWIEDZI RR2</b> |           |          |                                     |  |
| 1993                               | 83,2      | 2,55     | –                                   | -3,61%                                       |
| 1995                               | 81,0      | 2,85     | -1,3% <sup>a</sup>                  |  |
| 1997                               | 80,5      | 2,43     | -0,3% <sup>a</sup>                  |  |
| 1998                               | 79,1      | 1,97     | -1,8%                               |  |
| 2000                               | 73,3      | 2,05     | -3,7% <sup>a</sup>                  |  |
| 2001                               | 70,2      | 2,28     | -4,2%                               |  |
| 2002                               | 69,8      | 2,56     | -0,6%                               |  |
| 2003                               | 66,5      | 3,24     | -4,7%                               |  |
| 2004                               | 60,6      | 2,62     | -8,9%                               |  |
| 2005                               | 59,6      | 2,23     | -1,6%                               |  |
| 2006                               | 54,8      | 2,66     | -8,1%                               |  |
| 2007                               | 49,3      | 2,13     | -10,1%                              |  |
| 2008                               | 50,5      | 2,20     | 2,5%                                |  |
| 2009                               | 47,9      | 1,72     | -5,1%                               |  |
| 2010                               | 44,5      | 2,37     | -7,1%                               |  |
| <b>WSPÓLCZYNNIK KONTAKTÓW CON1</b> |           |          |                                     |  |
| 1993                               | 88,8      | 2,30     | –                                   | -2,15%                                       |
| 1995                               | 88,1      | 2,64     | -0,4% <sup>a</sup>                  |  |
| 1997                               | 89,4      | 2,00     | 0,7% <sup>a</sup>                   |  |
| 1998                               | 88,9      | 1,71     | -0,6%                               |  |
| 2000                               | 88,4      | 2,48     | -0,3% <sup>a</sup>                  |  |
| 2001                               | 87,6      | 1,31     | -0,9%                               |  |
| 2002                               | 86,0      | 2,25     | -1,8%                               |  |
| 2003                               | 81,7      | 6,71     | -5,0%                               |  |
| 2004                               | 71,4      | 3,18     | -12,7%                              |  |
| 2005                               | 71,1      | 2,54     | -0,4%                               |  |
| 2006                               | 66,7      | 2,22     | -6,2%                               |  |
| 2007                               | 62,6      | 2,93     | -6,1%                               |  |
| 2008                               | 64,4      | 1,91     | 2,9%                                |  |
| 2009                               | 63,8      | 1,57     | -1,0%                               |  |
| 2010                               | 61,4      | 2,28     | -3,7%                               |  |
| <b>WSPÓLCZYNNIK ODMÓW REF1</b>     |           |          |                                     |  |
| 1993                               | 5,0       | 0,89     | –                                   |  |
| 1995                               | 5,9       | 1,24     | 9,4% <sup>a</sup>                   |  |

| 1                                    | 2     | 3    | 4                  | 5      |
|--------------------------------------|-------|------|--------------------|--------|
| 1997                                 | 7,7   | 1,46 | 14,7% <sup>a</sup> | 6,55%  |
| 1998                                 | 8,9   | 1,33 | 15,7%              |        |
| 2000                                 | 14,1  | 1,34 | 29,6% <sup>a</sup> |        |
| 2001                                 | 16,5  | 1,77 | 16,5%              |        |
| 2002                                 | 15,1  | 2,22 | -8,2%              |        |
| 2003                                 | 15,1  | 6,44 | -0,1%              |        |
| 2004                                 | 8,6   | 0,84 | -42,8%             |        |
| 2005                                 | 9,2   | 0,62 | 6,2%               |        |
| 2006                                 | 10,0  | 1,00 | 9,5%               |        |
| 2007                                 | 10,4  | 1,44 | 3,9%               |        |
| 2008                                 | 12,0  | 1,66 | 14,7%              |        |
| 2009                                 | 13,6  | 1,19 | 13,8%              |        |
| 2010                                 | 14,7  | 1,34 | 7,8%               |        |
| <b>WSPÓLCZYNNIK KOOPERACJI COOP4</b> |       |      |                    |        |
| 1993                                 | 94,17 | 1,08 | –                  | -1,39% |
| 1995                                 | 92,98 | 1,38 | -0,6% <sup>a</sup> |        |
| 1997                                 | 91,10 | 1,66 | -1,0% <sup>a</sup> |        |
| 1998                                 | 89,61 | 1,46 | -1,6%              |        |
| 2000                                 | 83,30 | 1,23 | -3,5% <sup>a</sup> |        |
| 2001                                 | 80,50 | 2,00 | -3,4%              |        |
| 2002                                 | 81,87 | 2,48 | 1,7%               |        |
| 2003                                 | 81,67 | 6,57 | -0,2%              |        |
| 2004                                 | 87,03 | 1,11 | 6,6%               |        |
| 2005                                 | 86,19 | 0,82 | -1,0%              |        |
| 2006                                 | 83,74 | 1,90 | -2,8%              |        |
| 2007                                 | 80,39 | 1,92 | -4,0%              |        |
| 2008                                 | 79,77 | 2,58 | -0,8%              |        |
| 2009                                 | 76,88 | 1,72 | -3,6%              |        |
| 2010                                 | 74,18 | 2,23 | -3,5%              |        |

\*Na podstawie średniej geometrycznej (n = 18)

<sup>a</sup> Wielkość uśredniona

Źródło: opracowanie własne.

TABELA II. Wartości statystyk opisowych dla poszczególnych współczynników realizacji w klasach miejscowości

| Rok   | WIEŚ   | Miasto<br>do 19,9 tys. | Miasto<br>20–49,9 tys. | Miasto<br>50–99,9 tys. | Miasto<br>100–499,9 tys. | Miasto<br>pow.500 tys. |
|---|--------|------------------------|------------------------|------------------------|--------------------------|------------------------|
| 1   | 2      | 3                      | 4                      | 5                      | 6                        | 7                      |
| <b>WSPÓLCZYNNIK ODPOWIEDZI RR2</b>            |        |                        |                        |                        |                          |                        |
| Średnia arytmetyczna                          |        |                        |                        |                        |                          |                        |
| 1993  | 89,9   | 86,2                   | 81,9                   | 82,0                   | 77,3                     | 70,0                   |
| 1995  | 88,5   | 83,4                   | 79,7                   | 78,6                   | 75,5                     | 65,1                   |
| 1997  | 88,5   | 79,8                   | 79,8                   | 78,2                   | 75,8                     | 65,2                   |
| 1998  | 87,9   | 81,7                   | 79,0                   | 78,1                   | 72,0                     | 59,4                   |
| 2000  | 85,4   | 75,3                   | 72,8                   | 71,6                   | 64,5                     | 46,4                   |
| 2001  | 82,0   | 73,9                   | 66,4                   | 66,5                   | 63,5                     | 44,0                   |
| 2002  | 80,0   | 69,6                   | 66,6                   | 67,2                   | 64,6                     | 49,2                   |
| 2003  | 76,5   | 67,6                   | 62,6                   | 59,8                   | 63,9                     | 47,9                   |
| 2004  | 71,3   | 61,1                   | 56,0                   | 55,1                   | 53,0                     | 44,8                   |
| 2005  | 70,1   | 57,1                   | 56,5                   | 55,7                   | 52,7                     | 43,9                   |
| 2006  | 66,5   | 53,7                   | 50,5                   | 47,9                   | 47,2                     | 36,9                   |
| 2007  | 61,1   | 47,8                   | 42,4                   | 44,1                   | 40,8                     | 33,0                   |
| 2008  | 63,2   | 52,0                   | 45,3                   | 42,3                   | 39,9                     | 32,0                   |
| 2009  | 60,2   | 47,5                   | 42,4                   | 44,3                   | 37,5                     | 30,1                   |
| 2010  | 56,0   | 45,8                   | 39,0                   | 39,1                   | 34,4                     | 28,4                   |
| Średnie roczne tempo wzrostu(+) / spadku(-) * |        |                        |                        |                        |                          |                        |
|   | -2,75% | -3,65%                 | -4,28%                 | -4,26%                 | -4,66%                   | -5,16%                 |
| <b>WSPÓLCZYNNIK KONTAKTÓW CON1</b>            |        |                        |                        |                        |                          |                        |
| Średnia arytmetyczna                          |        |                        |                        |                        |                          |                        |
| 1993  | 95,0   | 91,9                   | 88,2                   | 89,4                   | 86,9                     | 85,9                   |
| 1995  | 93,9   | 90,8                   | 88,1                   | 88,5                   | 87,0                     | 82,3                   |
| 1997  | 94,3   | 89,9                   | 89,4                   | 88,2                   | 88,6                     | 84,6                   |
| 1998  | 94,2   | 90,6                   | 89,8                   | 90,1                   | 86,7                     | 83,4                   |
| 2000  | 94,0   | 90,1                   | 88,8                   | 87,5                   | 86,8                     | 81,5                   |
| 2001  | 92,8   | 89,0                   | 88,3                   | 87,0                   | 86,5                     | 81,8                   |
| 2002  | 91,5   | 86,8                   | 86,1                   | 86,4                   | 84,5                     | 81,7                   |
| 2003  | 85,8   | 80,9                   | 78,4                   | 77,2                   | 80,6                     | 74,2                   |
| 2004  | 78,8   | 71,8                   | 67,5                   | 66,3                   | 65,8                     | 62,2                   |
| 2005  | 77,6   | 67,7                   | 68,4                   | 68,7                   | 67,4                     | 62,7                   |
| 2006  | 74,1   | 64,2                   | 63,9                   | 60,6                   | 62,7                     | 57,1                   |
| 2007  | 70,5   | 60,5                   | 58,0                   | 58,8                   | 58,0                     | 51,6                   |
| 2008  | 73,1   | 64,4                   | 61,6                   | 59,2                   | 57,7                     | 51,6                   |
| 2009  | 71,9   | 62,7                   | 60,8                   | 61,2                   | 56,2                     | 53,4                   |
| 2010  | 68,8   | 61,4                   | 57,4                   | 57,1                   | 55,7                     | 51,6                   |
| Średnie roczne tempo wzrostu(+) / spadku(-) * |        |                        |                        |                        |                          |                        |
|   | -1,88% | -2,34%                 | -2,50%                 | -2,60%                 | -2,58%                   | -2,95%                 |

| 1   | 2      | 3      | 4      | 5      | 6      | 7      |
|---|--------|--------|--------|--------|--------|--------|
| <b>WSPÓLCZYNNIK ODMÓW REF1</b>                |        |        |        |        |        |        |
| Średnia arytmetyczna                          |        |        |        |        |        |        |
| 1993  | 2,3    | 3,6    | 4,4    | 4,3    | 7,6    | 12,4   |
| 1995  | 2,6    | 4,6    | 6,4    | 6,1    | 8,9    | 13,0   |
| 1997  | 3,6    | 7,9    | 7,5    | 7,7    | 10,6   | 16,3   |
| 1998  | 4,1    | 7,1    | 8,5    | 9,5    | 12,8   | 20,8   |
| 2000  | 6,6    | 12,6   | 13,9   | 14,7   | 20,1   | 31,3   |
| 2001  | 8,5    | 13,3   | 19,3   | 18,9   | 21,2   | 34,4   |
| 2002  | 8,5    | 14,7   | 17,7   | 17,3   | 17,8   | 29,2   |
| 2003  | 8,3    | 12,7   | 15,0   | 16,4   | 14,0   | 25,1   |
| 2004  | 5,4    | 8,3    | 9,5    | 10,0   | 10,6   | 15,1   |
| 2005  | 5,2    | 8,7    | 9,9    | 10,6   | 12,5   | 16,3   |
| 2006  | 5,9    | 9,2    | 11,6   | 10,9   | 13,3   | 18,3   |
| 2007  | 7,5    | 10,7   | 14,0   | 12,6   | 15,0   | 16,5   |
| 2008  | 7,8    | 10,7   | 14,0   | 14,7   | 15,6   | 17,9   |
| 2009  | 9,2    | 13,0   | 16,2   | 15,2   | 16,5   | 21,6   |
| 2010  | 10,5   | 13,4   | 16,1   | 15,8   | 19,1   | 21,4   |
| Średnie roczne tempo wzrostu(+) / spadku(-) * |        |        |        |        |        |        |
|   | 9,41%  | 8,10%  | 7,87%  | 7,91%  | 5,58%  | 3,26%  |
| <b>WSPÓLCZYNNIK KOOPERACJI COOP4</b>          |        |        |        |        |        |        |
| Średnia arytmetyczna                          |        |        |        |        |        |        |
| 1993  | 97,5   | 96,0   | 94,8   | 94,9   | 91,1   | 85,0   |
| 1995  | 97,1   | 94,7   | 92,5   | 93,1   | 89,4   | 83,4   |
| 1997  | 96,2   | 91,0   | 91,2   | 90,9   | 87,5   | 80,2   |
| 1998  | 95,6   | 92,1   | 90,2   | 89,4   | 84,9   | 74,1   |
| 2000  | 92,9   | 85,6   | 83,7   | 83,1   | 76,3   | 59,8   |
| 2001  | 90,6   | 84,8   | 77,4   | 77,9   | 75,0   | 56,3   |
| 2002  | 90,4   | 82,4   | 79,1   | 79,7   | 78,4   | 62,7   |
| 2003  | 90,2   | 84,2   | 80,7   | 78,5   | 82,0   | 64,6   |
| 2004  | 93,0   | 88,1   | 85,5   | 84,7   | 83,3   | 74,8   |
| 2005  | 93,1   | 86,7   | 85,1   | 84,0   | 80,9   | 73,0   |
| 2006  | 91,9   | 85,4   | 81,3   | 81,5   | 78,1   | 66,9   |
| 2007  | 89,1   | 81,8   | 75,1   | 77,8   | 73,1   | 66,7   |
| 2008  | 89,0   | 83,0   | 76,4   | 74,2   | 71,9   | 64,2   |
| 2009  | 86,8   | 78,5   | 72,3   | 74,5   | 69,5   | 58,3   |
| 2010  | 84,2   | 77,4   | 70,8   | 71,2   | 64,2   | 57,1   |
| Średnie roczne tempo wzrostu(+) / spadku(-) * |        |        |        |        |        |        |
|   | -0,86% | -1,26% | -1,70% | -1,68% | -2,03% | -2,31% |

\* Na podstawie średniej geometrycznej (n = 18).

Źródło: opracowanie własne.

TABELA 3. Wartości korelacji  $r$  Pearsona między współczynnikami realizacji w poszczególnych okresach

|   | RR2  | CON1   | REF1    | COOP4    |
|---|------|--------|---------|----------|
| Okres stosowania prób adresowych (1993-VIII 2003) |      |        |         |          |
| RR2   | 1,00 | 0,61** | -0,93** | 0,95**   |
| CON1  |      | 1,00   | -0,27*  | 0,34**   |
| REF1  |      |        | 1,00    | -0,997** |
| COOP4   |      |        |         | 1,00     |
| Okres stosowania prób imiennych (IX 2003-2011)    |      |        |         |          |
| RR2   | 1,00 | 0,95** | -0,85** | 0,94**   |
| CON1  |      | 1,00   | -0,65** | 0,79**   |
| REF1  |      |        | 1,00    | -0,98**  |
| COOP4   |      |        |         | 1,00     |
| Cały badany okres (1993-2011)                     |      |        |         |          |
| RR2   | 1,00 | 0,96** | -0,42** | 0,79**   |
| CON1  |      | 1,00   | -0,14   | 0,59**   |
| REF1  |      |        | 1,00    | -0,87**  |
| COOP4   |      |        |         | 1,00     |

\*\* $p < 0,01$ ; \* $p < 0,05$

Źródło: opracowanie własne.

*Katarzyna Grzeszkiewicz-Radulska*

**HOW IS THE NONRESPONSE PHENOMENON DEVELOPING IN POLAND?  
A TREND ANALYSIS OF NONRESPONSE IN POLAND IN 1993-2011  
BASED ON CBOS RESEARCH**

Abstract

Using AAPOR's final disposition codes, the author shows trends in nonresponse rates from 1993 to 2011 based on the personal interview survey (CAPI) "Current Problems and Events" conducted in Poland by the Public Opinion Research Center (CBOS). The author shows that since 1993 response rate (RR2) has remarkably decreased (by 40 percentage points) due to noncontacts and refusals. The findings of multivariate analysis demonstrate that the lower contact rate (CON1) can be explained by changes in sample design (i.e. using individual samples rather than household samples) and interviewer workload.

The transition of sampling choices however, positively influenced the level of respondents' cooperation (COOP4). The adverse factors affecting respondents' participation are: heavier interviewer workload, a negative social mood and an improvement in the economic situation.

**Key words:** nonresponse, contact rate, cooperation rate, individual samples, household samples, PAPI, CAPI

NATALIA HIPSZ\*  
Centrum Badania Opinii Społecznej

## STANDARYZOWANE TECHNIKI BADAWCZE JAKO SZCZEGÓLNY SPOSÓB KOMUNIKOWANIA SIĘ

### Streszczenie

Badania ilościowe w Polsce i na świecie realizowane są z wykorzystaniem technik badawczych zróżnicowanych w aspekcie warunków technologicznych i komunikacyjnych realizacji. Wysiłki naukowe podejmowane przez socjologów pozwalają przypuszczać, że zróżnicowane uwarunkowanie komunikacyjne pomiarów może przesądzać o braku ekwiwalentności, mniejszej bądź większej wiarygodności zgromadzonych danych. W niniejszym opracowaniu analizie poddano trzy standaryzowane techniki badawcze: osobisty wywiad kwestionariuszowy, telefoniczny wywiad kwestionariuszowy oraz ankietę audytoryjną. Artykuł, częściowo w drodze syntezy wyników prac empirycznych, dokonuje deskrypcji ilościowych podejść badawczych w kategoriach teorii komunikacji.

**Słowa kluczowe:** sytuacja komunikacyjna, pytania drażliwe, aprobata społeczna, wysiłek umysłowy, osobisty wywiad kwestionariuszowy, telefoniczny wywiad kwestionariuszowy, ankietę audytoryjną.

### WPROWADZENIE ]

Pojęcie technik badawczych odnosi się do procedury gromadzenia materiału empirycznego poddawanego analizie w toku badania socjologicznego. J. Lutyński pisze, iż „wszystkie techniki stanowią bardziej lub mniej rozbudowane kompleksy czynności i środków ujęte w reguły” [Lutyński 1983]. Sformułowanie to obejmuje swym zakresem zarówno wykorzystywane narzędzia badawcze, rolę badacza i jego współpracowników, jak i szerszy kontekst – również komunikacyjny – realizacji badania. Owe reguły determinują w myśl tego sposób zbiera-

\* n.hipsz@cbos.pl

nia danych, umożliwiając intersubiektywną komunikowalność i sprawdzalność wysiłków naukowych.

Zainteresowania badawcze artykułu koncentrują się na technikach standaryzowanych. Standaryzacja rozumiana jest tu jako ujednoczenie zdobytych informacji poprzez zastosowanie ściśle określonych wzorców w odniesieniu do czynności oraz środków badawczych. Standardy te to dyrektywy postępowania, identyczne we wszystkich sytuacjach badawczych – np. taki sam sposób formułowania pytań czy stosunek do respondenta. Jako kryterium podziału technik badawczych na standaryzowane i niestandaryzowane służy często fakt posługiwania się, bądź nie, pisemnymi formularzami, które w istotnym stopniu ograniczają swobodę, narzucają schematyzm i ukierunkowują akt pomiaru [por. Lutyński 1983]<sup>1</sup>. Wdrożenie pozornie nieskomplikowanej idei standaryzacji nastęrcza trudności w przypadku technik opartych na komunikowaniu się, w których może niekiedy dochodzić do przekłamań – ukrywania faktycznego stanu rzeczy – lub do niezrozumienia sformułowań zastosowanych w pytaniach narzędzia badawczego, szczególnie prawdopodobnego w badaniach zbiorowości niejednorodnych pod względem istotnych cech społeczno-demograficznych, np. wykształcenia [por. Fowler i in. 1990; Oppenheim 2004].

Jak zauważa P.B. Sztabiński, „realizacja badania jest prawie zawsze związana z tworzeniem pewnych sztucznych warunków, które umożliwiają jego przeprowadzenie, ale jednocześnie mogą mieć wpływ na uzyskane wyniki” [Sztabiński B. P. 1999b]. Wysiłki badawcze podejmowane przez socjologów niejednokrotnie dowodzą, iż nawet tak, mogło by się zdawać, subtelny, a nade wszystko bagatelizowany w codziennej praktyce badawczej czynnik, jak uwarunkowanie komunikacyjne pomiarów, może przesądzać o braku ekwiwalentności, mniejszej bądź większej wiarygodności zgromadzonych danych, uświadamiając tym samym, że sygnalizowany m.in. przez S. Nowaka problem „alienacji» metod badawczych” [por. Nowak 2006] nie kończy się na wyborze między ilościowym, a jakościowym podejściem do problematyki badawczej, wykazując jego aktualność w odniesieniu do procedur, których odmienność wydaje się zdecydowanie mniej intuicyjna, jak chociażby do całej gamy – cieszących się w dzisiejszych czasach niekwestionowaną popularnością – ilościowych technik badawczych o znacznym stopniu standaryzacji.

Niniejsze opracowanie analizuje trzy standaryzowane techniki badawcze – osobisty wywiad kwestionariuszowy, telefoniczny wywiad kwestionariuszowy

---

<sup>1</sup> Pojęcie „pomiaru” używane jest tu w znaczeniu wąskim i oznacza jednostkowy akt terenowej realizacji badania, tożsamy z przeprowadzaniem pojedynczego wywiadu czy ankiety.

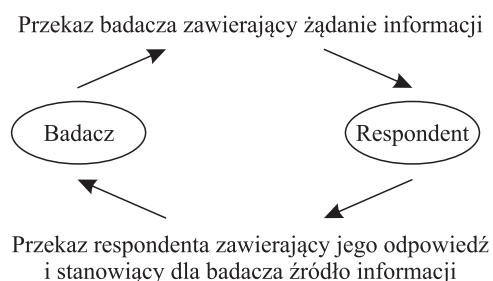


oraz ankietę audytoryjną – tworzące jeden z filarów, na których wspiera się proces gromadzenia wiedzy o współczesnym społeczeństwie [Lutyńska i in. 1983; Lutyński 1994; Słomczyński 2004], a przy tym zróżnicowane w aspektach kluczowych dla dywersyfikacji uwarunkowań komunikacyjnych pomiaru. W artykule, częściowo w drodze syntezy wyników prac empirycznych, dokonano deskrypcji ilościowych podejść badawczych w kategoriach teorii komunikacji społecznej.

### ISTOTA PROCESU WZAJEMNEGO KIEROWANEGO KOMUNIKOWANIA

Stosunek, w jakim pozostaje badacz i respondent w toku wywiadu kwestionariuszowego czy ankiety to wzajemnie kierowane komunikowanie. Istotę tego procesu stanowi zamknięty obieg informacji, w którym jedno ogniwo – badacz – emituje komunikat stanowiący żądanie informacji, natomiast drugie – respondent – skierowaną do badacza odpowiedź na to żądanie (patrz rysunek 1). Ten niezmienny w ramach sytuacji badawczej układ ról odróżnia sytuację standaryzowanych technik badawczych od zwykłej rozmowy, czyniąc ją szczególnym rodzajem komunikowania się [Lutyński 1983; zob. także Tourangeau 1988).

RYSUNEK 1. Schemat procesu wzajemnego kierowanego komunikowania w technikach otrzymywania materiałów



Źródło: [Lutyński 1983: 23].

Należy przy tym zaznaczyć, iż omawiany sposób komunikowania wymaga od uczestniczących w nim partnerów możliwie homogenicznego rozumienia kodów języka, w jakim generowany jest komunikat. Ponadto proces ten nie musi zachodzić w jednym czasie i miejscu oraz obejmuje zarówno przekazy ustne – angażujące zwykle także osobę ankietera – jak i pisemne [Lutyńska 1983; Lutyński 1983]. Z ostatnimi kwestiami wiąże się jeden z problemów wzajemnego kierowanego komunikowania – problem bezpośredniości i pośredniości pomiaru.

Zdekodowane żądanie informacji wywołuje u respondenta szereg procesów psychicznych, które określają jego reakcję w tym – co szczególnie istotne – treść udzielonej odpowiedzi. Z perspektywy badacza, w zależności od jej szczerości<sup>2</sup>, prawdziwości<sup>3</sup>, czy nawet faktu udzielenia, mogą to być zarówno procesy pożądane, jak i niepożądane [por. Lutyńska 1983; Lutyński 1983; Tourangeau i in. 2008]. Wśród czynników warunkujących przebieg procesów psychicznych towarzyszących definiowaniu sytuacji badania winno się wyróżnić szeroko pojęty kontekst sytuacyjny oraz wzory i modele społeczne, obyczajowe, czy kulturowe osoby badanej, w tym ukształtowane wcześniej postawy i wyobrażenia – patrz rysunek 2 [Bielawska, Słomczyński 1968; Lutyńska 1972, 1983; Lutyński 1983; Słomczyński 1983; Sołoma 1970].

Przez kontekst sytuacyjny należy rozumieć liczbę jednorazowo badanych osób, rodzaj dominujących w przekazie komunikatów werbalnych, zapośredniczenie (bądź jego brak) relacji badawczej przez medium komunikacyjne, jak również miejsce przeprowadzania pomiaru wpływające na to, z jaką rolą społeczną w danym momencie identyfikuje się respondent. Wymienione elementy w połączeniu z obecnością (bądź absencją) osób trzecich determinują stopień odczuwanej anonimowości i poufności badania [zob. także Gostkowski 1965; Hyman 1954], a nawet – szczególnie w odniesieniu do kwestii wspieranych przez instytucjonalny system kar i nagród – wpływają na pojawienie się odpowiedzi odzwierciedlających opinie publiczne: oficjalne, bądź prywatne [Lutyńska 1983; Lutyński 1983; Słomczyński 1966; Sułek 2002; zob. także Gostkowski 1966]. Duże znaczenie wydaje się tu także odgrywać pora dnia<sup>4</sup>, ogólne samopoczucie i nastroj respondent.

Kolejny czynnik obejmuje wzory i uwewnętrznione modele społeczne, obyczajowe i kulturowe, w tym wcześniej ukształtowane postawy i wyobrażenia. Pośród postaw największe znaczenie odgrywiają te odnoszące się bezpośrednio do instytucji reprezentowanej przez ankietera bądź sytuacji badania socjologicznego,

<sup>2</sup> Pod pojęciem „szczerości” rozumiana jest tu odpowiedź wrażliwa faktyczną postawę, wiedzę oraz gotowość do określonego zachowania się badanego będąca odpowiedzią subiektywnie prawdziwą – niezawierającą rezultatów świadomego wprowadzania w błąd badacza, czy ankietera; o szczerości odpowiedzi bliżej: [Nowak 1965: 61–64].

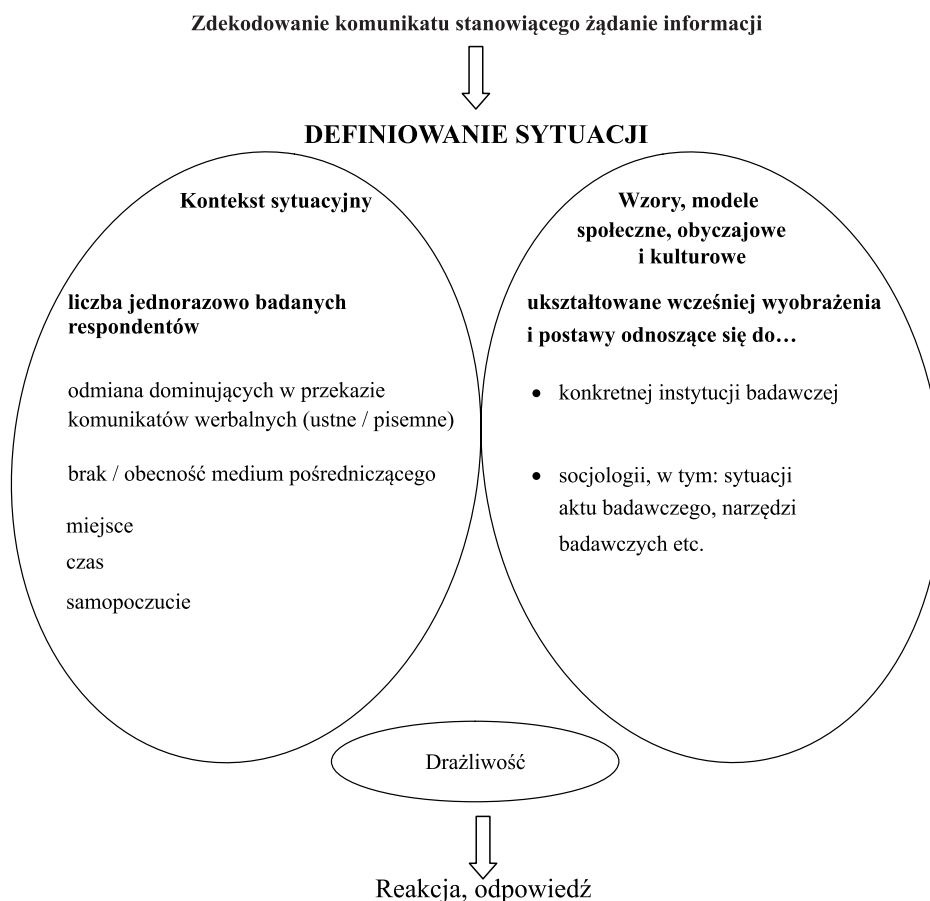
<sup>3</sup> Pod pojęciem „prawdziwości” rozumiana jest tu odpowiedź niebędąca rezultatem – nie-uświadomionego respondentowi – stwarzania fikcji o sobie. Owe fikcje – generowane w wyniku „osłuchania się” z często głoszonymi w danej zbiorowości postulatami – sprawiają, iż odpowiedź na postawiane pytanie nie odzwierciedla rzeczywistej postawy respondenta; o stwarzaniu fikcji o sobie bliżej: [Ossowski 1967: 23].

<sup>4</sup> Często spotykana w praktyce badawczej późna pora aranżacji wywiadów telefonicznych bywa odbierana jako zachowanie niegrzeczne.

ewentualnie – zwykle w przypadku braku odpowiednich doświadczeń – innych kontekstów silnie je imitujących. Reakcja na kontakt z ankierem albo badaczem nie jest prostą wypadkową uwarunkowań sytuacyjnych. Respondenci mogą mieć pewne wyobrażenia o celu badań, obawy lub oczekiwania wynikające z podzielanych wartości oraz wykształconego dotąd stosunku do narzędzi badawczych czy do socjologii jako takiej. Ogólne nastawienie do badań socjologicznych determinowane jest często stopniem akulturacji technik badawczych oraz internalizacji elementów roli respondenta w grupach, do których należą ankietowani [Gostkowski 1983; Lutyński 1983].

Niejako na pograniczu obydwu tych czynników wyróżnić można niezmiernie istotny, jeśli nie najważniejszy, bodziec – drażliwość. W odniesieniu do ilościowych technik badawczych pod owym pojęciem kryje się szereg negatywnych odczuć (wstydu, lęku, gniewu, zażenowania czy złości), za których wystąpienie odpowiadać może zarówno problematyka poszczególnych pytań – przykładowo stosujących niezrozumiałe terminy, dotyczących kwestii intymnych, sformułowanych w sposób osobisty (nie zaś projekcyjnie), poruszających zagadnienia regulowane przez opinię publiczną bądź demaskujących niewiedzę respondenta – jak i ogólny charakter poszczególnych faz lub całego aktu badawczego – brak jasności celów projektu, nienaturalne zmienianie tematu, niedostatek poczucia konfidențialności procesu albo ogólna trudność z akceptacją konwencji wzajemnego kierowanego komunikowania. Definiowana w ten sposób drażliwość ma charakter subiektywny: percepcja drażliwości pytania lub całego badania zależy zarówno od czynników sytuacyjnych (np. stopnia bezpośredniości przekazu w danej technice badawczej), jak i od norm zachowań, wiedzy oraz reakcji ugruntowanych w kapitale kulturowym, obyczajowości jednostki czy szerszej zbiorowości [Bielawska, Słomczyński 1968; Kubiak i in. 1992; Sułek 2002; zob. także Lutyńska 2000].

RYSUNEK 2. Czynniki determinujące procesy psychiczne w toku definiowania sytuacji badania przez respondenta



Źródło: opracowanie własne.

Charakter procesów zachodzących w psychice respondenta – determinowany zmiennymi uwarunkowaniami sytuacyjnymi oraz bardziej utrwalonymi (znajdującymi wyraz w konkretnych postawach i wyobrażeniach) wzorami i modelami – odciska swoje piętno na całym akcie badawczym. Różne konfiguracje tych elementów decydują o efektach nieustannej definicji sytuacji, która nie jest dla badanego łatwa [Rostocki 1999]. Jak trafnie zauważają A. Kubiak, I. Przybyłowska oraz W. A. Rostocki, „neutralność stosowanego w socjologii terminu respondent

sprawia, że nie wszyscy socjologowie pamiętają, że respondent to przede wszystkim osoba badana. Co więcej, osoba badana w warunkach i na zasadach, które nie ona ustala i których do końca wywiadu nie zawsze jest świadoma” [Kubiak i in. 1992]. W konsekwencji, z punktu widzenia ankietowanego, realizacja wywiadu ilościowego – szczególnie zaś wykorzystującego ustne komunikaty werbalne oraz pełną gamę komunikatów niewerbalnych – to permanentne podejmowanie próby wejścia w rolę respondenta i sprostania jej wymogom. Problematyczność tego procesu może skutkować trudnościami napotkanymi przez badacza w trakcie uzyskiwania zgody na udział w projekcie, jego przerwaniem, nieprawidłowym udziałem w badaniu, bądź – przede wszystkim w technikach ankietowych – brakami danych [Kubiak i in. 1992; Sztabiński F. 1992].

Odmowa wzięcia udziału w badaniu może być konsekwencją niedostatecznego prestiżu prowadzącej je instytucji, drażliwej problematyki projektu, braku wiary w jego anonimowość bądź poufność, braku czasu, czy wreszcie negatywnego odbioru osoby ankietera – jego cech socjodemograficznych, osobowościowych oraz kulturowych. P.B. Sztabiński posiłkując się wynikami badań zrealizowanych przez I. Stoop, wyróżnia trzy zasadnicze grupy czynników wpływających na decyzję potencjalnego respondenta o uczestniczeniu (bądź nieuczestniczeniu) w sondażu: ogólny stosunek do badań ilościowych – poziom zaufania do ich wyników oraz traktowanie udziału w tego typu przedsięwzięciach jako interesującej formy spędzenia czasu; stosunek do konkretnego projektu badawczego – postrzegana ważność sondażu; „substantywne” charakterystyki badania – społeczną doniosłość (uczestnictwo w projektach realizowanych przez agendy rządowe lub naukowe bywa postrzegane jako społeczny obowiązek) oraz osobistą ważność i atrakcyjność jego problematyki, sponsoring [Stoop 2005; Sztabiński P.B. 2006].

Do przzerwania pomiaru doprowadza brak akceptacji dla jego konwencji, zmęczenie i zniecierpliwienie procesem oraz problemy natury intelektualnej – chociażby niezrozumienie terminów stosowanych w pytaniach [Kubiak i in. 1992; Sztabiński F. 1992].

Nieprawidłowy udział w badaniu odnosi się do zagadnienia prawdziwości i wiarygodności odpowiedzi. Respondenci mogą bowiem odpowiadać na pytania ukrywając rzeczywiste fakty i opinie lub unikając wysiłku – bez refleksji nad poruszonymi kwestiami. Artefakty, a więc odpowiedzi wykreowane przez sytuację badania, stanowią znaczne zagrożenie dla rzetelności zebranych danych empirycznych.

Obecność braków danych (wyjątkowo częsta w formularzach wypełnianych samodzielnie przez ankietowanych) jest natomiast konsekwencją nieudzielenia odpowiedzi na jedno bądź wiele pytań kwestionariusza, które może wynikać

z drażliwości narzędzia badawczego, jak również z ogólnego braku akceptacji dla nienaturalnej sytuacji społecznej wykreowanej przez proces badawczy [Bielawska, Słomczyński 1968].

Odmienność i nieadekwatność wzajemnego kierowanego komunikowania do doświadczeń znanych z codziennej interakcji wywołuje dyskomfort również u ankietera, który prowadząc wywiad ilościowy musi sprostać licznym, niewspółmiernym do codziennego doświadczenia, restrykcyjnym wymogom standaryzacji, przy jednoczesnej konieczności podtrzymania możliwie satysfakcjonującego kontaktu z respondentem. Spowodowane tym napięcie emocjonalne może stanowić argument do rezygnacji z sumiennego stosowania zaleceń szkolenia ankieterskiego [Kubiak i in. 1992].

Poczynione dotąd rozważania uświadamiają każdorazową unikatowość aktu badawczego, którego przebieg i efekty w znacznym stopniu zależą jednak od specyfiki sytuacji, także komunikacyjnej, realizacji.

### **OSOBISTY WYWIAD KWESTIONARIUSZOWY**

Osobisty wywiad kwestionariuszowy to bodaj najbardziej klasyczna ilościowa technika badawcza. Tym, co odróżnia ową procedurę od pozostałych standaryzowanych technik jest bezpośredniość sytuacji komunikacyjnej – wywiad osobisty przybiera postać szczególnej rozmowy typu face-to-face, wykorzystującej zarówno ustne komunikaty werbalne, jak i pełną gamę komunikatów niewerbalnych<sup>5</sup>; osobisty, indywidualny charakter – w procesie, obok badacza bądź ankietera, bierze udział tylko jeden respondent; wspomaganie kwestionariuszem wypełnianym przez badacza lub ankietera – w ujęciu wzorcowym ankietowany nie ma wglądu do pytań formularza [Sztabiński F. 1992].

Sytuacja wywiadu implikuje niedogodności zarówno po stronie wykonującego swą pracę ankietera, jak i biorącego udział w badaniu respondenta. Ankieter jest z punktu widzenia badanego obcą osobą, narzucającą mu silnie sformalizowaną, jednostronną interakcję, która sytuując respondenta na pozycji podporządkowanej może zostać odebrana jako instrumentalne, a wręcz manipulatorskie traktowanie jego osoby. Wywiad opiera się na nienaturalnej strukturze rozmowy, w której – w zwykłe zamkniętych, przybliżających charakter przesłuchania oraz eksponujących odrębność od bliskiego i znanego ankietowanemu świata społecznego pytaniach – stosuje się w znacznej mierze nieadekwatne

<sup>5</sup> Niekiedy także pisemne komunikaty werbalne – w postaci załączonych do narzędzia badawczego kart dla respondenta.

do codziennego doświadczenia terminy (obcość kulturowa pytań). Dodatkowo badany narażony jest na ujawnienie braku kompetencji czy niewiedzy [Kubiak i in. 1992]. Wszystko to sprawia, iż – jak piszą A. Kubiak, I. Przybyłowska oraz W. A. Rostocki – wywiad kwestionariuszowy jest „sytuacją społecznie otwartą na rozmaite podejrzenia respondenta dotyczące sensu i istoty spotkania z ankierem (obcym), celu rozmowy (niejasnego) i wynikających z jej odbycia konsekwencji (nieznanych)” [Kubiak i in. 1992].

Negatywne odczucia potęguje bezpośredni charakter ustnej i naocznej sytuacji komunikacyjnej wywiadu osobistego. Bezpośredniość procesu może podawać w wątpliwość przekonanie o poufności i anonimowości aktu badawczego [Słomczyński 1983]. Interaktywność badania wzmaga ponadto dyskomfort odczuwany przez ankietowanego w sytuacji udzielania odpowiedzi na pytanie podchwytliwe, osobiste lub poruszające kwestie poniekąd regulowane przez opinię publiczną [Oppenheim 2004]. Respondenci niejednokrotnie mogą czuć się zawstydzeni lub zażenowani zainteresowaniami badawczymi projektu, a niechęci otwarcia się przed osobą obcą towarzyszy – częstokroć nieuświadomiony – wewnętrzny przymus udzielania odpowiedzi zgodnych z przyjętymi w danej zbiorowości normami społecznymi i kulturowym, potęgowany przez dążenie do przedstawienia siebie w jak najlepszym świetle (ang. *social desirability*; kwestie te są często podejmowane w obcojęzycznej literaturze przedmiotu, por. np. Crowne i in. 1964; Edwards 1957, 1990). Jak stwierdza H. Hyman, swoiste społeczne uwikłanie respondenta przesądza o tym, iż sytuację wywiadu traktuje on nie tylko w wymiarze *stricte* zadaniowym, lecz również w społecznym. Efektem tego stanu rzeczy jest – szczególnie silna w sytuacji bezpośredniego kontaktu z ankierem – tendencja do udzielania odpowiedzi akceptowalnych z punktu widzenia ogólnospołecznych norm [Hyman 1954; zob. także DeMaio 1984; Sztabiński F. 1999a; Lee 1993]<sup>6</sup>, mogąca prowadzić do swoistych przekłamań w zgromadzonym materiale empirycznym, zmniejszając tym samym jego wartość poznawczą [Lutyńska 1972; Tangney, Fischer 1995; Tangney i in. 1996]. Ów problem dostrzega A. Sułek pisząc, iż sytuacja wywiadu jest *de facto* sytuacją publiczną, co skłania respondentów do występowania w nim w roli oficjalnej: ukrywania wartości odczuwanych – opinii prywatnych, a deklarowania powszechnie uznawanych – opinii oficjalnych [Gostkowski 1966; Ossowski 1967; Sułek 1990].

<sup>6</sup> Warto jednakże zauważyć, iż, wedle wyników Polskiego Generalnego Sondażu Społecznego 1994 oraz Europejskiego Sondażu Społecznego 2004, coraz mniej Polaków uważa, iż wyniki badań służą sprawdzeniu, czy poglądy ludzi są „właściwe”; por. [Sztabiński P. B. 2006: 22].

Ponadto w bezpośredniej relacji badany częściej niż w kontaktach bardziej impersonalnych decyduje się na podejmowanie działań niejako wbrew sobie, kosztem własnego komfortu psychicznego. Do takich zachowań należy zaliczyć kontynuowanie wywiadu pomimo zniecierpliwienia i znudzenia, spowodowane problematycznością przerwania interakcji, związaną z silnym oddziaływaniem kulturowo sankcjonowanej reguły grzeczności. Rezultatem nadgorliwego respektowania owej normy jest „wymuszone” uciekanie do odpowiedzi typu „nie wiem”, „trudno powiedzieć” bądź udzielanie odpowiedzi nieprzemyślanych, a w efekcie nieprawidłowość aktu badawczego [Kubiak i in. 1992; Sztabiński F. 1992].

Zaletą bezpośredniości relacji z ankierem jest natomiast skłonność osób badanych do intensywnego wysiłku umysłowego skutkującego – zarówno w przypadku wywiadów silnie standaryzowanych, jak i wykraczających w swą konstrukcji poza pytania zamknięte – pełniejszymi i bardziej przemyślanymi odpowiedziami [Sztabiński P.B. 1999a].

Jak wykazują badania, aby zwiększyć wiarygodność uzyskanych odpowiedzi osobisty wywiad kwestionariuszowy winien być przeprowadzany w miejscu najmniej narażonym na obecność osób trzecich, których nawet pośredni udział w procesie badawczym może skutkować „zanieczyszczeniem” zebranego materiału empirycznego [Lutyńska 1972]. Nie ulega kwestii, iż największa szansa sprawowania kontroli nad tym czynnikiem jest niezaprzeczalnym atutem wywiadu osobistego.

### TELEFONICZNY WYWIAD KWESTIONARIUSZOWY

Wywiady telefoniczne zagościły w praktyce badawczej na początku lat 70. XX w. [Sztabiński P. B. 1999a]. Popularyzacja owej techniki stała się możliwa w chwili telefonizacji znacznej części społeczeństwa – nie tylko specjalnych jego kategorii, których dominacja w próbie może prowadzić do jej stronniczości [Frankfort-Nachmias, Nachmias 2001]<sup>7</sup>.

O specyfice techniki telefonicznego wywiadu kwestionariuszowego stanowi jej częściowo zapośredniczony charakter – wywiad ten jest realizowany z wykorzystaniem medium oraz ustnych komunikatów werbalnych przy praktycznym deficycie komunikatów niewerbalnych (z wyjątkiem chronemiki, będącej funkcją zachowań głosowych); potencjalną obecnością osób trzecich w toku realizacji badania – badacz czy ankieter nie mają wpływu na sytuację, w jakiej „zastaną”

<sup>7</sup> O słabościach wywiadów telefonicznych bliżej [Klecka, Tuchfarber 1978: 105–114]; zob. także [Jabkowski 2007].



respondenta; wspomaganie kwestionariuszem wypełnianym przez badacza lub ankietera – podobnie jak w wywiadzie osobistym respondent nie ma wglądu do pytań formularza, o ile jednak w relacji *face-to-face* jest to zwykle niepoddające się weryfikacji życzenie, w interakcji zapośredniczonej przez telefon staje się standardem.

Niedogodności, z jakimi borykają się uwikłani w realizację wywiadu telefonicznego ankieter i respondent są zbliżone do sygnalizowanych wcześniej ogólnych trudności ilościowych wywiadów kwestionariuszowych (wszystkich technik angażujących w postępowanie badawcze ankietera).

Tym, co najwyraźniej odróżnia sytuację wywiadu telefonicznego od osobistego jest pośredni charakter relacji ankieter–respondent. Okoliczność ta skutkuje reinterpretacją problemów znamienych dla naocznej sytuacji wywiadu osobistego oraz wyłonieniem się nowych, determinowanych obecnością medium komunikacyjnego.

Pośredniość procesu decyduje o znacznie silniejszym aniżeli w przypadku kontaktu *face-to-face* poczuciu anonimowości, czy poufności badania [Sztabiński F., Sztabiński P.B. 1997]. Zapośredniczenie interakcji skutkuje mniejszą niż w wywiadzie osobistym skłonnością do udzielania odpowiedzi pod presją powszechnie przyjmowanych norm oraz wartości – odpowiedzi społecznie akceptowanych. Jak wykazują badania H. Crossley i H. Parry, dzieje się tak przede wszystkim wówczas, gdy pytanie dotyczy kwestii o większej społecznej drażliwości. Brak fizycznej obecności ankietera prowadzi zdaniem autorów do swoistej depersonalizacji powodującej słabsze uwikłanie społeczne, a w efekcie ograniczenie tendencji do przemyślenia w odpowiedziach opinii publicznych jedynie celem zachowania twarzy i kreowania pozytywnego wizerunku własnej osoby [Crossley, Parry 1950; Sztabiński P.B. 1999b; zob. także Groves 1989; Oppenheim 2004; Hoinville, Sykes 1985].

Inną konsekwencją pośredniości kontaktu z ankieterem jest mniej zintensyfikowany wysiłek umysłowy (ang. *satisficing*, por. Green i in. 2001; Krosnick 1991), przejawiający się w mniejszej precyzyjności odpowiedzi<sup>8</sup>, a także krótszym czasie poświęcanym na generowanie komunikatu zwrotnego w reakcji na zadane pytania [Sztabiński P.B. 1999a, 1999b; Sztabiński F., Sztabiński P.B. 1997]. Mniejszy wysiłek intelektualny oraz większa szybkość wywiadu telefonicznego przesądza o tym, iż respondenci w nim uczestniczący częściej, aniżeli badani w sposób bezpośredni udzielają odpowiedzi skrajnych i schematycznych – co

---

<sup>8</sup> O precyzyjności odpowiedzi należy tu mówić przede wszystkim w przypadku pytań otwartych.

uwidacznia się chociażby w operowaniu całkowitymi lub „okrągłymi” liczbami w pytaniach stosujących skalę – oraz konfirmatywnych – wskazują odpowiedź „tak”, bądź „zgadzam się” [Presser, Schuman 1996; Sztabiński P.B. 1999a]. Schematyzm jest według J. Beavin, D. Jacksona i P. Watzlawick’a, rezultatem ograniczenia komunikatów niewerbalnych odpowiedzialnych za przekazywanie większości znaczeń społecznie istotnych – respondent nie ma możliwości dokładnego zastanowienia się nad odpowiedzią, bowiem niemal niewykonalne okazuje się zakomunikowanie tego momentu ankietarowi (niesygnalizowane w sposób lingwistyczny zastanawianie się doprowadza do niezręcznej i trudnej do zinterpretowania w relacji niebezpośredniej ciszy) – na rzecz ustnych komunikatów werbalnych, których funkcją jest przede wszystkim przekazywanie „suchych” informacji (por. Beavin i in. 1967). O znacznej szybkości wywiadu przesądza ponadto tendencja do przerywania ankietarowi zanim zdola on zaprezentować wszystkie odpowiedzi, czego rezultatem może być dominacja wskazań rozpoczynających kafelety nad tymi, które ją zamykają.

Kolejną własnością wywiadu telefonicznego jest rodzący się między ankietarzem, a respondentem dystans psychiczny, którego konsekwencją może być mniej kłopotliwe niż w przypadku relacji *face-to-face* przerwanie wywiadu – brak wzrokowego kontaktu z rozmówcą sprawia, iż odłożenie słuchawki wiąże się z niewielkimi tylko pejoratywnymi konotacjami psychologicznymi [Głodowski 2006; Frankfort-Nachmias, Nachmias 2001].

O szczególnym charakterze wywiadu telefonicznego przesądza nie tylko pośredniość relacji badanego i ankietarza. Inną znamioną jego cechą jest brak kontroli badacza nad miejscem, okolicznościami w jakich – po stronie respondenta – realizowany jest wywiad. Badanie J.M. Fridion poświęcone zachowaniom seksualnym Francuzów wykazuje, iż obecność w trakcie przeprowadzania pomiaru małżonka ankietowanej za pośrednictwem telefonu respondentki lub małżonki respondenta modyfikuje odpowiedzi na pytania o kwestie drażliwe – w omawianym projekcie chociażby o liczbę partnerów seksualnych [Sztabiński F., Sztabiński P.B. 1997]. Obecność osób trzecich może skutkować uzyskaniem informacji nieszczerych czy poniekąd kolektywnych – utrata kontroli nad sytuacją, w jakiej ankietowany odpowiada na pytania kwestionariusza, wiąże się z ryzykiem aktywnego włączenia się w ów proces jego otoczenia [Sztabiński P.B. 1999b]. Ponadto można spodziewać się narastania dystansu psychicznego między respondentem i ankietarzem, wynikającego z bardziej „uspołecznionego” charakteru bezpośrednich relacji osoby badanej z otaczającymi ją, z reguły dobrze znanymi, ludźmi.

Niemожność decydowania o miejscu rozmowy ankietowanego – znamienna w dobie dominacji telefonii komórkowej nad klasyczną stacjonarną wersją medium – pociąga za sobą znaczny wzrost prawdopodobieństwa modyfikacji odpowiedzi nie tylko za sprawą wpływu osób trzecich, lecz również wyglądu, bądź ról społecznych, w jakich badany na co dzień w nim występuje [Lutyńska 1972].

### ANKIETA AUDYTORYJNA

Technika ankiety audytoryjnej wykorzystywana jest w badaniach czyniących za obiekt dociekań uformowane zbiorowości; swoiste całości społeczne [Bielawska, Słomczyński 1968].

Ankieta audytoryjna odznacza się pośredniością przekazu – charakterystyczny dla omawianych dotąd technik przekaz ustny zostaje tu zastąpiony pisemnym, co przesądza o zdominowaniu całego aktu badawczego werbalnymi komunikatami pisemnymi; kolektywnym charakterem – kwestionariusze, po uprzedniej aranżacji, uzupełniane są przez większą grupę osób zgromadzonych w tym samym miejscu i czasie [Gostkowski 1983]; wykorzystywaniem narzędzia badawczego opatrzonego stosowną instrukcją i wypełnianym samodzielnie przez respondenta – osoba badana ma możliwość zapoznania się z całym kwestionariuszem ankiety przed udzieleniem odpowiedzi na jakiegokolwiek jej pytanie [Bielawska, Słomczyński 1968; Lutyński 1977].

Z perspektywy procesów komunikacyjnych postępowanie badawcze w ankiecie audytoryjnej zasadniczo różni się od procedur wywiadu. O owej różnicy przesądza między innymi pisemny charakter przekazu, pociągający za sobą zarówno pozytywne, jak i negatywne konsekwencje dla efektywności komunikowania. Można oczekiwać, że fakt samodzielnego notowania odpowiedzi na spisane w formularzu pytania oraz brak ustnej komunikacji między badaczem i respondentem wpływa na zwiększenie poczucia anonimowości i poufności przekazywanych informacji. Komfort psychiczny wzmacnia zapewne często spotykana praktyka umieszczania wypełnionych kwestionariuszy w urnie. Poczucie konfidențialności winno skutkować redukcją presji wywieranej na osobie badanej przez potrzebę aprobaty społecznej<sup>9</sup> i – w efekcie – większą aniżeli w technikach wywiadu skłonnością do udzielania odpowiedzi zgodnych z opiniami prywatnymi.

---

<sup>9</sup> Autorka ma tu na myśli wymiar osobisty – inną kwestią wydaje się próba wywarcia pozytywnego wrażenia respondenta jako członka „grupy” i wiążąca się z nią anonimowość zbiorowa; por. [Hyman 1954].

Podobnie jak w innych sytuacjach badawczych, istnieje ryzyko, iż opinie prywatne zostaną nieintencjonalnie zdeformowane poprzez stwarzane o sobie fikcje [Sołoma 1970], natomiast duża konfidencjonalność procesu może niejednokrotnie skutkować rezygnacją z nadmiernego wysiłku umysłowego, a nawet – przysparzającym licznym trudności interpretacyjnych i niemalże pozbawionym negatywnych konsekwencji psychologicznych u respondenta – nieudzieleniem odpowiedzi na pytania ankiety. Brak odpowiedzi wynika na ogół z przeświadczenia, że sam w sobie jest on odpowiedzią, z przeoczenia polecenia, bądź świadomego działania osoby badanej. Celowe nieudzielenie odpowiedzi może być podyktowane niezrozumieniem treści pytania, a w przypadku narzędzi mniej standaryzowanych, także nieumiejętnością sformułowania własnej opinii lub przekazania posiadanej wiedzy. Istotną barierę stanowi tutaj ograniczony kontakt ankietowanych z badaczem. Jak piszą B. Bielawska i K.M. Słomczyński, „badacz posiada [...] wprawdzie możliwość wyjaśnienia wątpliwości, jakie nasuwa respondentom sam kwestionariusz, ale nie jest to środek, który gwarantowałby wysoki stopień kontroli nad jego rozumieniem” [Bielawska, Słomczyński 1968]. Do powstrzymania się od odpowiedzi przyczynia się również brak wyrobionej opinii lub wiedzy bądź niechęć do ich ujawnienia – ów problem dotyczy przede wszystkim pytań o wysokim stopniu drażliwości oraz osób nieprzekonanych do procedur szeroko rozumianej socjologii.

Inną cechą odróżniającą technikę ankiety audytoryjnej od porównywanych tu technik wywiadu jest zbiorowy charakter badania. Kolektywizm w połączeniu z redukcją ustnej komunikacji werbalnej sprawia, iż sytuacja badawcza ankiety ma silnie standaryzowany charakter, co eliminuje, a przynajmniej ogranicza, prawdopodobieństwo wystąpienia w toku pomiaru wielu błędów o charakterze systematycznym (choćby efektu ankietarskiego).

Zmniejszenie kontroli nad wypełnianiem kwestionariuszy może przyczyniać się do wzmożonej komunikacji między respondentami – ryzyko to potęguje fakt, iż w przypadku ankiety audytoryjnej nie są oni sobie obcy, a brak kontaktu *face-to-face* z badaczem nie wpływa pozytywnie, szczególnie w przypadku pytań trudnych, na intensyfikację wysiłku umysłowego respondentów. W rezultacie badani mogą udzielać odpowiedzi z perspektywy roli pełnionej w badanej zbiorowości lub wyrażać opinie poniekąd kolektywne.

Jak trafnie zauważa J. Lutyński, „z różnymi technikami wywiadu łączy ankietę to, że respondenci udzielają w niej informacji, odpowiadając na różne pytania, z którymi jednak w ankiecie, w przeciwieństwie do wywiadu, respondent może się zazwyczaj zapoznać od razu w całości, przeczytawszy cały kwestionariusz przed udzieleniem odpowiedzi na jakiegokolwiek pytanie” [Lutyński 1977].

W niektórych sytuacjach okoliczność ta, generując tak zwany efekt kontekstu [zob. także Daniłowicz 1999], może znacznie zaburzać ideę standaryzacji – przykładowo kolejność odczytywania wariantów kafeterii odpowiedzi; w innych – chociażby w przypadku konwencji spotykanej często w badaniach marketingowych, gdzie pytanie zamknięte zostaje poprzedzone pytaniem otwartym na ten sam temat – z uwagi na wysoki poziom sugestywności, winna zasadniczo wykluczać zastosowanie omawianej techniki w procesie gromadzenia materiału empirycznego.

### ZAKOŃCZENIE

Nierównoważność procesów wzajemnego kierowanego komunikowania wynika przede wszystkim ze stopnia zapośredniczenia relacji badawczej oraz, ściśle z tym związane, charakteru dekodowania komunikatu i definiowania sytuacji badania przez respondenta. Bezpośredniość – przekaz ustny, brak obecności medium – oraz pośredniość – przekaz pisemny, obecność medium – poszczególnych technik badawczych, pociąga za sobą daleko idące konsekwencje metodologiczne.

Bezpośredni, ustny charakter wywiadu osobistego przesądza o większym wzajemnym oddziaływaniu zaangażowanych w pomiar osób – wpływ przekazu jest tu wzmacniany, lub modyfikowany przez komunikaty niewerbalne. Ponadto odczytujący respondentowi pytania ankieter, ma w wywiadzie osobistym – dzięki interakcji *face-to-face* – relatywnie dużą kontrolę nad procesem pomiaru. Inaczej przedstawia się sytuacja ankiety audytoryjnej i wywiadu telefonicznego. W technice ankiety pisemny przekaz stanowi dominujący bodziec komunikacyjny, co ogranicza do minimum możliwość wywierania wzajemnego wpływu przez badacza i osobę badaną, zmniejszając jednocześnie szanse kontroli całego procesu. Z kolei kontekst realizacji wywiadu telefonicznego warunkuje – za sprawą pośrednictwa medium ograniczającego znacznie komunikację nielingwistyczną – słabsze niż w wywiadzie osobistym i – dzięki dominacji ustnych komunikatów werbalnych – silniejsze aniżeli w ankiecie audytoryjnej wzajemne oddziaływanie zaangażowanych w pomiar osób oraz – ponownie w konsekwencji zapośredniczenia ich relacji przez medium komunikacyjne – równie niski jak w technice ankiety stopień kontroli aktu badawczego (patrz tabele 1 i 2).

TABELA 1. Użycie komunikatów werbalnych i niewerbalnych przez badacza/ankietera i respondenta w poszczególnych standaryzowanych technikach badawczych

|                                       | Komunikaty werbalne |          |           |          | Komunikaty niewerbalne |           |
|---------------------------------------|---------------------|----------|-----------|----------|------------------------|-----------|
|                                       | Ustne               |          | Pisemne   |          | B                      | R         |
|                                       | B                   | R        | B         | R        |                        |           |
| Osobisty wywiad kwestionariuszowy     | +++<br>++           | ++<br>++ | +         | -        | +++<br>++              | +++<br>++ |
| Telefoniczny wywiad kwestionariuszowy | +++<br>++           | ++<br>++ | -         | -        | +                      | +         |
| Ankieta audytoryjna                   | ++                  | +        | +++<br>++ | ++<br>++ | ++                     | ++        |

B – badacz, R – respondent, znak „+” oznacza występowanie komunikatów danego rodzaju (im więcej powtórzeń, tym większy ich udział w procesie komunikowania się), natomiast znak „-” – brak komunikatów danego rodzaju

Źródło: opracowanie własne.

TABELA 2. Charakterystyka kontekstu komunikacyjnego poszczególnych ilościowych technik badawczych

| Rodzaj kontaktu badacza/ankietera z respondentem ze względu na: | Osobisty wywiad kwestionariuszowy | Telefoniczny wywiad kwestionariuszowy | Ankieta audytoryjna |
|---|-----------------------------------|---------------------------------------|---------------------|
| liczbę jednorazowo badanych respondentów                        | Indywidualny                      | Indywidualny                          | Kolektywny          |
| brak/ obecność medium pośredniczącego                           | Bezpośredni                       | Pośredni                              | Bezpośredni         |
| odmianę dominujących w przekazy komunikatów werbalnych          | Ustne (bezpośrednie)              | Ustne (bezpośrednie)                  | Pisemne (pośrednie) |

Źródło: opracowanie własne.

Można więc przypuszczać, iż technika osobistego wywiadu kwestionariuszowego – za sprawą silnego społecznego wymiaru oraz wysokiego stopnia dwuwymiarowej bezpośredniości – pociąga za sobą najbardziej odczuwalne dla respondenta konsekwencje psychologiczne. Wykorzystujący medium ustny przekaz wywiadu telefonicznego – średnie, zaś stwarzająca poczucie anonimowości pisemna ankieta audytoryjna – najsłabsze. Z drugiej jednak strony, zważywszy na postępujące wraz z upośrednieniem relacji badawczej swoiste „rozleniwianie” respondenta, należy spodziewać się większej intensyfikacji wysiłku umysłowego, a więc uzyskiwania bardziej przemyślanych, mniej schematycznych odpowiedzi w – przyjmowanym za najbardziej drażliwy – interpersonalnym kontekście badawczym.

**BIBLIOGRAFIA**

- Beavin J., Jackson D., Watzlawick P. [1967], *Pragmatics of human communication*, New York: W. W. Norton.
- Bielawska B., Słomczyński K.M. [1968], *Z doświadczeń realizacji ankiety audytoryjnej i wywiadu zbiorowego w środowisku studenckim*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Ossolineum.
- Crossley H.M., Parry H. J. [1950], *Validity of responses to survey questions*, "Public Opinion Quarterly", vol. 14.
- Crowne D.P., Marlowe D. [1964], *The approval motive*, New York: Wiley Publishing.
- Daniłowicz P. [1999], *Analiza efektu kontekstu w interpretacji danych surveyowych: Koncepcja Norberta Schwarza*, „Przegląd Socjologiczny”, t. XLVIII/1.
- DeMaio Th. J. [1984], *Social desirability and survey measurement: A review*, [w:] Ch.E. Turner, E. Martin (eds), *Surveying subjective phenomena*, vol. 2, New York: Russel Sage Foundation.
- Edwards A.L. [1957], *The social desirability variable in personality assessment and research*, New York: Dryden.
- Edwards A.L. [1990], *Construct validity and social desirability*, "American Psychologist", vol. 45.
- Fowler F.J., Mangione Th. W. [1990], *Standardized survey interviewing: Minimizing interviewer-related error*, London: Sage.
- Frankfort-Nachmias Ch., Nachmias D. [2001], *Metody badawcze w naukach społecznych*, Poznań: Zysk i S-ka.
- Głodowski W. [2006], *Komunikowanie interpersonalne*, Warszawa: Studio Emka.
- Gostkowski Z. [1965], *Badania adekwatności technik badawczych w krajach na drodze rozwoju*, „Przegląd Socjologiczny”, t. XIX/2.
- Gostkowski Z. (red.) [1966], *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław: Ossolineum.
- Gostkowski Z. [1983], *Sytuacja społeczna wywiadu kwestionariuszowego: Społeczno-kulturowe uwarunkowanie wywiadu kwestionariuszowego: O założeniach i potrzebie empirycznych badań nad wywiadem kwestionariuszowym*, [w:] K. Lutyńska, A.P. Wejland (red.), *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*, Wrocław: Ossolineum.
- Gostkowski Z. (red.) [1990], *Analizy i próby technik badawczych w socjologii*, t. 8, Wrocław: Ossolineum.
- Gostkowski Z. (red.) [1992], *Analizy i próby technik badawczych w socjologii*, t. 9, Wrocław: Ossolineum.
- Gostkowski Z., Lutyński J. (red.) [1968], *Analizy i próby technik badawczych w socjologii*, t. 2, Wrocław: Ossolineum.
- Gostkowski Z., Lutyński J. (red.) [1970], *Analizy i próby technik badawczych w socjologii*, t. 3, Wrocław: Ossolineum.
- Gostkowski Z., Lutyński J. (red.) [1972], *Analizy i próby technik badawczych w socjologii*, t. 4, Wrocław: Ossolineum.
- Green M.C., Krosnick J.A., Holbrook A.L. [2001], *The survey response process in telephone and face-to-face surveys: Differences in respondent satisficing and social desirability response bias*, b.m.w.
- Groves R.M. [1989], *Survey errors and survey costs*, New York: Wiley-Interscience.

- Hoinville G., Sykes W. [1985], *Telephone interviewing on a survey of social attitudes: A comparison with face-to-face procedures*, London: Social and Community Planning Research.
- Hyman H. [1954], *Interviewing in social research*, Chicago: University of Chicago Press.
- Jabkowski P. [2007], *Wpływ niezrealizowania części wywiadów na trafność wnioskowania statystycznego w badaniach społecznych: Technika wywiadu kwestionariuszowego oraz telefonicznego w świetle błędów nielosowych*, „ASK: Społeczeństwo. Badania. Metody”, vol. 16.
- Klecka W.R., Tuchfarber A.J. [1978], *Random Digital dialing: A comparison to personal survey*, “Public Opinion Quarterly”, vol. 42.
- Krosnick J.A. [1991], *Response strategies for coping with the cognitive demands of attitude measures in surveys*, “Applied Cognitive Psychology”, vol. 5.
- Kubiak A., Przybyłowska I., Rostocki W.A. [1992], *Społeczna przestrzeń wywiadu kwestionariuszowego*, [w:] Z. Gostkowski (red.), *Analizy i próby technik badawczych w socjologii*, t. 9, Wrocław: Ossolineum.
- Lee R.M. [1993], *Doing research on sensitive topics*, London: Sage.
- Lutyńska K. [1972], *Miejsce przeprowadzania wywiadu i jego oddziaływanie*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 4, Wrocław: Ossolineum.
- Lutyńska K. [1983], *Sytuacja społeczna wywiadu kwestionariuszowego: Wprowadzenie*, [w:] K. Lutyńska, A.P. Wejland (red.), *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*. Wrocław: Ossolineum.
- Lutyńska K. [1997], *Wpływ ankierski w pierwszej fazie badań kwestionariuszowych*, „ASK: Społeczeństwo. Badania. Metody”, vol. 1–2 (5–6).
- Lutyńska K. [2000], *Problemy drażliwe w badaniach socjologicznych w Polsce*, „ASK: Społeczeństwo. Badania. Metody”, vol. 9.
- Lutyńska K., Wejland A.P. (red.) [1983], *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*, Wrocław: Ossolineum.
- Lutyński J. [1977], *Ankieta i jej rodzaje*, Warszawa: WSNS.
- Lutyński J. [1983], *Wywiad kwestionariuszowy jako technika otrzymywania danych: Techniki otrzymywania materiałów i ich podział*, [w:] K. Lutyńska, A.P. Wejland (red.), *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*, Wrocław: Ossolineum.
- Lutyński J. [1994], *Metody badań społecznych: Wybrane zagadnienia*, Łódź: Łódzkie Towarzystwo Naukowe.
- Nowak S. [2006], *Metodologia badań społecznych*, Warszawa: PWN.
- Nowak S. (red.) [1965], *Metody badań socjologicznych*, Warszawa: PWN.
- Oppenheim A. V. [2004], *Kwestionariusze, wywiady, pomiary postaw*, Poznań: Zysk i S-ka.
- Ossowski S. [1967], *Z zagadnień psychologii społecznej*, Warszawa: PWN.
- Presser S., Schuman H. [1996], *Questions and answers in attitude measurement*, „Psychological Bulletin”, vol. 103, no. 3. surveys, London: Sage.
- Rostocki W.A. [1999], *Socjologia a prywatność*, „Przegląd Socjologiczny”, t. XLVIII/1.
- Słomczyński K.M. [1966], *Wpływ oficjalnej i prywatnej sytuacji wywiadu na wypowiedzi respondentów*, [w:] Z. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław: Ossolineum.
- Słomczyński K.M. [1983], *Sytuacja społeczna wywiadu kwestionariuszowego: Społeczno-kulturowe uwarunkowanie wywiadu kwestionariuszowego. O założeniach i potrzebie empirycznych badań nad wywiadem kwestionariuszowym*, [w:] K. Lutyńska, A.P. Wejland (red.), *Wywiad kwestionariuszowy: Analizy teoretyczne i badania empiryczne*, Wrocław: Ossolineum.



- Słomczyński K.M. [2004], *Polska metodologia socjologicznych badań empirycznych na tle międzynarodowym*, [w:] P.B. Sztabiński, F. Sztabiński, Z. Sawiński (red.), *Nowe metody, nowe podejścia badawcze w naukach społecznych*, Warszawa: IFiS PAN.
- Sołoma L. [1970], *Analiza porównawcza wartości danych zdobytych przy pomocy różnych technik*, [w:] W. Gostkowski, J. Lutyński (red.), *Analizy i próby technik badawczych w socjologii*, t. 3, Wrocław: Ossolineum.
- Stoop I.A.L. [2005], *The hunt for the last respondent: Nonresponse in sample surveys*, Hague: Social and Cultural Planning Office.
- Sułek A. [1990], *W terenie, w archiwum i w laboratorium: Studia nad warsztatem socjologa*, Warszawa: Wydawnictwo IS UW.
- Sułek A. [2002], *Ogród metodologii socjologicznej*, Warszawa: Scholar.
- Sztabiński F. [1992], *Wywiad kwestionariuszowy jako technika badawcza*, [w:] P. Daniłowicz, Z. Sawiński, F. Sztabiński, P.B. Sztabiński, A.P. Wejland, *Podręcznik socjologicznych badań ankietowych: Ankieter w procesie badawczym*, Warszawa: IFiS PAN.
- Sztabiński F., Sztabiński P.B. [1997], *Wartość odpowiedzi w wywiadzie telefonicznym ze wspomaganie komputerowym (CATI)*, „ASK: Społeczeństwo. Badania. Metody”, vol. 1–2 (5–6).
- Sztabiński P.B. [1999a], *Wywiad telefoniczny ze wspomaganie komputerowym (CATI) a wywiad kwestionariuszowy: Problem porównywalności wyników*, [w:] H. Domański, K. Lutyńska, A. W. Rostocki (red.), *Spojrzenie na metodę: Studia z metodologii badań socjologicznych*, Warszawa: IFiS PAN.
- Sztabiński P. B. [1999b], *Wywiad telefoniczny ze wspomaganie komputerowym: Nowa technika badawcza czy wywiad kwestionariuszowy przez telefon?*, „Przegląd Socjologiczny”, t. XLVIII/1.
- Sztabiński P.B. [2006], *Dlaczego respondenci uczestniczą lub nie uczestniczą w badaniach? Porównanie 1999–2004*, „ASK: Społeczeństwo. Badania. Metody”, vol. 15.
- Tangney J.P., Fischer K. [1995], *Self-conscious emotions: shame, guilt, embarrassment, and pride*, New York: Guilford Press.
- Tangney J.P., Miller R.W., Flicker L., Barlow D.H. [1996], *Are shame, guilt and embarrassment distinct emotions?*, „Journal of Personality and Social Psychology”, vol. 70.
- Tourangeau R., Rasinski K. [1988], *Cognitive process underlying context effect in attitude*
- Tourangeau R., Rips L.J., Rasinski K. [2008], *The psychology of survey response*, New York: Cambridge University Press.

Natalia Hipsz

**STANDARDIZED RESEARCH TECHNIQUES AS A SPECIAL TYPE  
OF COMMUNICATION – THE COMPARISON OF THE PERSONAL STANDARDIZED  
INTERVIEW, THE TELEPHONE INTERVIEW AND THE AUDITORIUM SURVEY**

Abstract

Quantitative researches which are conducted in Poland and all over the world use techniques varied in terms of technological and communication conditions of realization. Scientific efforts which have been made by sociologists suggest that the differential communication conditioning of

measurement may prejudice of nonequivalence, higher or lower level of reliability of gathered data. The paper analyzes three standardized research techniques: the personal standardized interview, the telephone interview and the auditorium survey. The article, partly by synthesis of empirical studies' results, makes a description of quantitative research methods in the context of the theory of communication.

**Key words:** communication context, sensitive survey questions, mental effort, social desirability, personal standardized interview, telephone interview, auditorium survey

WOJCIECH JABŁOŃSKI\*  
Uniwersytet Łódzki

## SYTUACJE TRUDNE W WYWIADZIE TELEFONICZNYM

### Streszczenie

W artykule przedstawiam wyniki badania metodologicznego, jakie przeprowadziłem wśród ankierów telefonicznych współpracujących z czołowymi polskimi agencjami badawczymi. W projekcie starałem się ustalić, jak bardzo różne czynności wykonywane w studiu CATI są dla ankierów stresujące. Ponadto, zależało mi na identyfikacji strategii, jakie stosują ankierzy, gdy napotykają na kłopotliwych respondentów.

Wyniki badania pokazują, że stresujące dla ankierów są dwa rodzaje zadań: czynności związane z ryzykiem przerwania wywiadu przez respondenta oraz czynności powodujące poznawcze obciążenie ankiera. Strategie wdrażane przez ankierów w trakcie rozmowy z trudnym respondentem zależą od rodzaju kłopotliwej sytuacji. W przypadku, gdy respondent ma trudności ze zrozumieniem pytań, większość ankierów destandardyzuje przebieg wywiadu, upraszczając słownictwo i składnię. Reakcje ankierów w innych sytuacjach (respondenci niechętni, zirytowani, cechujący się dużą aktywnością uboczną) są bardziej skomplikowane i zależą od szeregu czynników.

**Słowa kluczowe:** CATI, ankierzy, badania telefoniczne, standaryzacja

Wywiad telefoniczny ze wspomaganie komputerowym (CATI – *Computer Assisted Telephone Interview*) jest obecnie w Polsce jedną z najpopularniejszych technik zbierania danych w komercyjnych badaniach opinii i rynku. W 2010 r. osoby przebadane przy użyciu tej techniki stanowiły 33,4% wszystkich respondentów biorących udział w badaniach marketingowych, co daje ok. 2,3 miliona

---

\* jablonex@poczta.fm

przeprowadzonych wywiadów telefonicznych [Chojnowski, Wódkowski 2011]. W wielu krajach Europy Zachodniej i w USA wywiad CATI (oprócz bardzo szerokiego wykorzystania w badaniach rynkowych) znajduje też regularne zastosowanie w projektach prowadzonych w ramach statystyki publicznej<sup>1</sup> oraz w socjologicznych badaniach akademickich.

W artykule dokonuję krótkiej charakterystyki wywiadu telefonicznego, porównując go z wywiadem *face-to-face*; skupiam się w szczególności na specyfice komunikacji telefonicznej i wynikających z niej trudnościach, które mogą towarzyszyć ankieterowi w trakcie prowadzenia aranżacji wywiadu oraz zadawania pytań i rejestrowania odpowiedzi respondentów. W dalszej części artykułu przedstawiam wyniki własnego badania metodologicznego, jakie zrealizowałem wśród ankieterów CATI współpracujących z największymi polskimi agencjami świadczącymi usługi w zakresie badań rynku i opinii<sup>2</sup>.

### CHARAKTERYSTYKA TECHNIKI

Popularność wywiadu telefonicznego związana jest przede wszystkim z krótkim (w porównaniu do wywiadów terenowych) okresem realizacji, gwarantowanym elektroniczną formą narzędzia badawczego i scentralizowanym procesem zbierania danych oraz relatywnie niskim kosztem przeprowadzenia pojedynczego wywiadu.

W wywiadzie CATI tradycyjny papierowy kwestionariusz jest zastąpiony skrypcem elektronicznym, który wyświetla na ekranie komputera ankieterskiego pytania przewidziane do zadania respondentowi; rejestruje on również odpowiedzi badanych (ankieterzy wprowadzają je bezpośrednio do komputera). Technologia taka eliminuje konieczność druku formularzy, ich dystrybucji (dostarczania ankieterom i późniejszego odbierania wypełnionych kwestionariuszy), jak również potrzebę ręcznego tworzenia komputerowego zbioru danych – kodowania/wprowadzania danych.

<sup>1</sup> Warto wspomnieć, że podczas Narodowego Spisu Powszechnego Ludności i Mieszkań 2011 jedną z technik zbierania danych (obok wywiadu osobistego, ankiety internetowej i źródeł administracyjnych) był właśnie wywiad telefoniczny. Choć wykorzystany w niewielkim zakresie (w kanale CATI w badaniu pełnym spisano jedynie 2,9% respondentów, a w badaniu reprezentacyjnym 1% próby [Biuletyn... 2011]), wpisze się być może do repertuaru technik wykorzystywanych przez polską statystykę publiczną.

<sup>2</sup> Badania, których wyniki przedstawiam w artykule, przeprowadziłem w ramach przygotowywania rozprawy doktorskiej pt. *Wywiad telefoniczny ze wspomaganie komputerowym (CATI) ze szczególnym uwzględnieniem działań ankieterskich*.

Elektroniczna forma narzędzia w wywiadzie CATI zapewnia też automatyczną kontrolę spójności danych (na przykład system zablokuje próbę wstawienia kodu spoza zakresu zmiennej, nie „przeoczy” pytania, które ma być odczytane, itd.), jak również daje możliwość implementacji do skryptu rozwiązań metodologicznych niedostępnych w tradycyjnym wywiadzie kwestionariuszowym. Do powszechnie wykorzystywanych należą na przykład rotacja i randomizacja wyświetlania pytań lub kategorii odpowiedzi, emisja nagrań, nagrywanie fragmentów wywiadu<sup>3</sup> [zob. Sawiński 1996].

Choć realizacja projektów CATI wymaga poniesienia sporych wydatków wyjściowych (zakup sprzętu komputerowego, oprogramowania, wyposażenia studia, itp.) [Kelly i in. 2008], koszt przeprowadzenia pojedynczego wywiadu telefonicznego jest znacząco niższy w porównaniu do wywiadu opierającego się na bezpośrednim kontakcie ankietera z respondentem. Szacuje się [Hołdakowski 2006], że w warunkach polskich standardowy wywiad CATI jest około 2–3 razy niższy aniżeli wywiad *face-to-face*. Różnica w kosztach wzrasta w przypadku dużego terytorialnego rozproszenia próby oraz w badaniach *business-to-business*, zakładających zwykle rekrutację respondentów trudnodostępnych.

Przygotowanie i realizacja badania z użyciem techniki CATI różni się całkowicie od procedur stosowanych w przypadku badań terenowych. Po pierwsze, skomplikowaniu ulega proces doboru próby i kontroli jej statystycznej reprezentatywności. W projektach zakładających estymację wyników na populacje ogólnonarodowe najczęściej niemożliwe jest losowanie respondentów z operatów – wykazy zawierające wyczerpujące informacje o numerach telefonów „przypisanych” do jednostek nie istnieją. Wyjątek stanowią tutaj m.in. kraje skandynawskie, w których tego typu spisy funkcjonują [Kalsbeek, Agans 2008] – wykonalne jest więc tam losowanie próby na takiej samej zasadzie, jak ma to miejsce w badaniach terenowych, a następnie dzwonienie na numery, jakimi – według zapisów w bazie – posługują się respondenci. Najczęściej jednak konieczne jest korzystanie z procedury losowego generowania numerów (RDD – *Random Digit Dialing*), polegającej – w dużym skrócie – na „tworzeniu” przez komputer numeru telefonicznego, z którym będzie próbował połączyć się ankieter. Operacja ta przebiega z uwzględnieniem specyfiki funkcjonującej na danym obszarze infrastruktury telekomunikacyjnej (format numeru, zasada konstrukcji

<sup>3</sup> Opiswane w tym akapicie cechy wywiadu CATI właściwe są też dla innych technik, których realizacja jest wspomagana komputerowo, na przykład dla wywiadu CAPI (*Computer Assisted Personal Interview* – wywiad *face-to-face* z laptopem), czy też badań internetowych – CAWI (*Computer Assisted Web Interview*).

prefiksów, logika przydzielania przez operatorów numerów abonentom, itp.) [Sawiński 1996; Steeh 2008].

Co więcej, w badaniach CATI niezbędne jest zwykle łączenie w próbach numerów stacjonarnych i komórkowych – właściwa dla większości krajów spadająca penetracja telefonii stacjonarnej i rosnąca liczba abonentów telefonii komórkowej sprawia, iż próby generalne ludności bazujące jedynie na użytkownikach telefonów „domowych” straciłyby walor reprezentatywności [Keeter i in. 2007]. W Polsce, według szacunków [Kujawski i in. 2008], w 2008 r. telefon stacjonarny posiadało jedynie 47% gospodarstw domowych, natomiast z telefonu komórkowego korzystało 75% Polaków.

Pod drugie, w badaniach CATI stosuje się odmienną od tej właściwej dla badań terenowych organizację pracy sieci ankietarskiej. Ankieterzy zgromadzeni są w jednym bądź kilku *call center* (każde studio wyposażone jest w odpowiednią liczbę stanowisk komputerowych – od kilkunastu do kilkuset), skąd telefonują do respondentów; praca ankietników zorganizowana jest w kilkugodzinne sesje. Znajduje to odzwierciedlenie w stosowanych przez ośrodki badawcze procedurach *human resources*, przede wszystkim tych dotyczących wynagradzania i kontroli pracy ankietników. Ankieterzy są najczęściej opłacani za przepracowaną godzinę (a nie za zrealizowany przypadek, jak ma to miejsce w pracy ankietników terenowych), zaś kontrola ich działań odbywa się w czasie rzeczywistym – nadzorujący pracę studia superwizor widzi pracującego ankietera, może włączyć podsłuch wywiadu, który akurat jest realizowany, ma też możliwość podglądu ekranu ankietarskiego (widzi, jakie odpowiedzi zaznacza ankietnik) [Steve i in. 2008; *Polskie standardy...* 2010]. Świadomość intensywności działań kontrolnych wymusza na ankietnikach ściślejsze (aniżeli w przypadku wywiadów *face-to-face*<sup>4</sup>) przestrzeganie zaleceń instrukcji, a co za tym idzie umożliwia osiągnięcie znacznie większej standaryzacji zachowania ankietników [Lavrakas 1993; Fowler, Mangione 1990].

Po trzecie, charakterystyczny dla wywiadu CATI jest brak fizycznej obecności ankietera w trakcie przeprowadzania badania – proces komunikowania się między ankietnikiem a respondentem przebiega więc z wyłączeniem kanału niewerbalnego – jedynie na płaszczyźnie werbalnej i parajęzykowej. Odmienność kontaktu

<sup>4</sup> W badaniach *face-to-face* bieżący monitoring pracy ankietników jest niemożliwy – stosuje się w nich wywiad kontrolny (lub kontrolną ankietę pocztową) zawierające pytania dotyczące samego wywiadu (jak długo trwał, czego dotyczył, itp.), jak również powtórzone niektóre z pytań zadanych wcześniej przez ankietera [*Polskie standardy...* 2010; Sztabiński 1995]. Więcej możliwości kontrolnych (aczkolwiek wciąż mniej niż wywiad CATI) daje wywiad osobisty ze wspomaganie komputerowym – CAPI [zob. Sawiński 2005].

w wywiadzie osobistym i wywiadzie CATI ma wpływ na uzyskiwane odpowiedzi, jak również na przebieg interakcji na linii ankieter-respondent [de Leeuw 2002].

Z jednej strony, stosunkowo duży dystans między ankieterem a respondentem w wywiadzie CATI działa korzystnie, ograniczając rozmiar efektu ankieterskiego. Jak pokazują badania [np. Tucker 1983; Groves, Magilavy 1986], zniekształcający wpływ cech ankietera na odpowiedzi respondentów jest trochę mniejszy w wywiadzie telefonicznym aniżeli *face-to-face*<sup>5</sup>. Co więcej, również społeczny wymiar wywiadu CATI wydaje się być mniej istotny, co sprawia, że społeczne uwikłanie respondenta – charakterystyczne dla wywiadu osobistego [zob. Hyman 1954; Kubiak i in. 1992] – jest tutaj osłabione. W efekcie w wywiadzie telefonicznym obserwujemy często słabszą skłonność respondentów do udzielania odpowiedzi społecznie akceptowanych, to jest takich, które są zgodne z ogólnie przyjętymi normami [Sztabiński 1999; Frey 1989; Kormendi 1988].

Z drugiej strony, wywiad telefoniczny cechuje mniejsza (w porównaniu do wywiadu osobistego) skuteczność zabiegów aranżacyjnych. Brak kontaktu wzrokowego sprawia, iż ankieter otrzymuje mało wskazówek mogących wzmocnić perswazyjność jego komunikatów (ankieter nie widzi respondenta, otoczenia, w jakim on się znajduje – mieszkania, okolicy, itp.) [Couper, Groves 2002], a jedynym pozawerbalnym narzędziem wpływu na badanego jest umiejętne posługiwanie się głosem (intonacja, tempo, częstotliwość, zabarwienie emocjonalne, itp.) [Groves i in. 2008; Oksenberg, Cannell 1988]. Kontakt telefoniczny łatwiej też jest respondentowi przerwać; rozłączenie rozmowy jest „prostsze” niż zamknięcie przed ankieterem drzwi. Co więcej, w wywiadzie *face-to-face*, zakładając, że uda się go rozpocząć, zakończenie rozmowy w trakcie i wyproszenie ankietera z mieszkania jest stosunkowo trudne do przeprowadzenia – w wywiadzie CATI wystarczy przerwanie połączenia [Groves 1990]. Poza tym, ankieter telefoniczny nie ma żadnych możliwości kontrolowania warunków, w jakich realizowany jest wywiad, zwłaszcza ograniczania liczby czynności, które w tym samym czasie wykonuje respondent. W wywiadzie CATI, w szczególności, jeśli dzwoni się na numery komórkowe, respondent zwykle robi coś innego: gotuje obiad, uczestniczy w spotkaniu, prowadzi samochód itp. [Holbrook i in. 2003].

Przeprowadzając wywiady telefoniczne trzeba też – co oczywiste – liczyć się z bardzo ograniczonymi możliwościami stosowania bodźców wizualnych [de Leeuw 2002]. Użycie kart respondenta, typowe dla wywiadu osobistego, stanowi tutaj problem, chyba że karty te zostaną respondentowi dostarczone

---

<sup>5</sup> Oprócz braku fizycznej obecności ankietera czynnikiem zmniejszającym wpływ ankieterski w wywiadzie CATI jest też, wspomniana wcześniej, silna standaryzacja zachowania ankieterów.

wcześniej (mailem lub w przesyłce pocztowej) lub też jednym z zadań badanego jest przygotowanie takich kart (ankieter prosi o przygotowanie kartki, długopisu i poleca – na przykład – narysować skalę). Najczęściej brak materiałów wizualnych rekompensuje się specjalnymi rozwiązaniami stosowanymi w konstrukcji pytań. Chodzi tutaj przede wszystkim o zastępowanie skal semantycznych skalami numerycznymi lub też „rozbijanie” (*unfolding*) [Groves 1979] pytań ze skalami semantycznymi. Polega ono na zadawaniu w pierwszej kolejności pytania bardziej ogólnego (na przykład, czy respondent jest z czegoś zadowolony czy niezadowolony), a później pytania szczegółowego, którego treść jest uzależniona od wcześniejszej odpowiedzi (np. czy badany jest bardzo zadowolony czy trochę zadowolony). Do innych rozwiązań należy stosowanie technik wyobrażeniowych (na przykład „proszę sobie wyobrazić drabinę/termometr”), stosowanie pytań prekategoryzowanych (tj. otwartych dla respondenta, ale zamkniętych dla ankietera) oraz generalne ograniczanie liczby kategorii odpowiedzi w pytaniach oraz skracanie samych odpowiedzi [Frey 1989; de Leeuw 2002].

Kolejne problemy związane są ze specyfiką rozmowy telefonicznej – w porównaniu do wywiadu osobistego charakteryzują ją szybsze tempo mówienia [Groves 1990] oraz większa skłonność do unikania pauz [de Leeuw 2002]. Cechy te mają konsekwencje dla uzyskiwanych odpowiedzi. Charakterystyczne między innymi dla wywiadu CATI zachowania respondenckie Krosnick [1999] określa mianem *satisficing*. Chodzi tutaj o takie reakcje badanych, które mają na celu ograniczenie wysiłku poznawczego potrzebnego do udzielenia rzetelnej odpowiedzi na pytanie ankietera. Przykładami *satisficing* są: „potakiwanie”, „zgadzanie się” ze stwierdzeniami odczytywanymi przez ankietera niezależnie od ich treści (*acquiescence*); wielokrotne posługiwanie się tą samą opcją w odniesieniu do różnych itemów pytania, gdy kategorie odpowiedzi pozostają niezmiennie oraz wybieranie opcji „trudno powiedzieć”, jeżeli jest ona proponowana przez ankietera [Krosnick 1999; Holbrook i in. 2003]. Ponadto, analizy metodologiczne pokazują, że w wywiadzie telefonicznym uzyskuje się mniej wyczerpujące odpowiedzi respondentów na pytania otwarte [Sykes, Collins 1988], występuje też silniejsza skłonność badanych do wybierania na skalach opcji skrajnych a w pytaniach z odczytywaną listą kategorii – odpowiedzi zaprezentowanych jako pierwsze [Jordan i in. 1980; Groves 1979]<sup>6</sup>.

<sup>6</sup> Choć trudno jest pisać o generalnych prawidłowościach, przegląd literatury przedmiotu zdaje się sugerować, że prawdopodobieństwo pojawienia się opisywanych trudności jest większe, gdy respondentami są osoby starsze i/lub o niskim poziomie wykształcenia [Loosveldt 1997; Herzog, Rodgers 1988] oraz gdy wywiad jest długi [Holbrook i in. 2003; Frey 1989].



### ANKIETERZY CATI O SWOJEJ PRACY

Choć w literaturze przedmiotu (m.in. tej już cytowanej; poza tym: Groves i in. [2009]; Biemer i Lyberg [2003]) spotkać można wiele analiz dotyczących wpływu rozmaitych charakterystyk ankietera CATI (jego płci, wieku, przekonań, itp.) na jakość uzyskiwanych danych, mało jest przedstawionych wyników badań odnoszących się do opinii ankieterów na temat ich pracy – trudności, z jakimi się spotykają i sposobów, jakie stosują, aby je pokonać.

Tarnai i Paxson [2005] dokonali podsumowania wyników analiz metodologicznych, jakie były prowadzone w trakcie 12 sondaży telefonicznych. W odniesieniu do każdego zrealizowanego przypadku (było ich łącznie prawie 25 tysięcy) proszono ankieterów o ocenę jakości wywiadu oraz – jeżeli ocena nie była najwyższa – o wskazanie przyczyn wyjaśniających problem. Jedynie 10% wywiadów uznano za niedoskonałe. Wśród powodów, jakie mogły do tego prowadzić ankieterzy najczęściej wskazywali na trudności ze zrozumieniem pytania przez respondenta (21%), kłopoty ze słuchem (11%), zakłócenia pochodzące z otoczenia respondenta (9%), niedostateczną znajomość języka angielskiego (9%) i niedołążność, chorobę respondenta (7%). Stwierdzono zależność między oceną ankieterską a liczbą braków danych w wywiadzie – im mniej było braków, tym lepszą notę otrzymywał wywiad. Autorzy wyrażają pogląd, że opinie ankieterów o wywiadzie są dobrym wskaźnikiem jakości rozmowy i w wielu przypadkach pozwalają na identyfikację źródeł problemów, z jakimi zmagają się respondenci. Na przykład, w opisywanym badaniu dość duża frekwencja przypadków nierozumienia przez badanych pytań może świadczyć o konieczności wprowadzenia zmian w skrypcie wywiadu albo też osłabienia stopnia standaryzacji rozmowy [Conrad, Schober 2000].

Fazzi i in. [2009] opisują wyniki badania jakościowego, jakie przeprowadzono z 26 ankieterami CATI zatrudnionymi we włoskich agencjach realizujących wywiady telefoniczne. W indywidualnych wywiadach swobodnych pytano ankieterów o ich doświadczenia związane z pracą w studiu – w szczególności o to, jak często i w jakim zakresie zdarza im się postępować wbrew regule standaryzacji. Badani przywoływali wiele sytuacji, w których nie przestrzegali ujednolicających zaleceń instrukcji. Robili to, by uprzyjemnić wywiad, uczynić go mniej formalnym, bardziej zbliżonym do normalnej rozmowy lub też – choć rzadziej – by zmniejszyć ryzyko przerwania wywiadu przez respondenta. Autorzy zwracają uwagę na nieprzystawalność zaleceń obecnych w większości podręczników dla ankieterów do realiów pracy w studiu CATI; postulują wzbogacenie programu treningu ankieterskiego o elementy związane z radzeniem sobie z trudnymi sytuacjami, jakie mogą się pojawić w trakcie rozmowy.

Badanie metodologiczne, którego wyniki przedstawię w dalszej części tekstu, miało na celu znalezienie odpowiedzi na dwa pytania problemowe. Bazując na relacjach samych ankierów starałem się – po pierwsze – ustalić, jak bardzo różne czynności wykonywane w studiu CATI są dla ankierów stresujące. Po drugie, chciałem dokonać rekonstrukcji strategii, jakie stosują ankierzy, kiedy napotykają na „trudnego” respondenta, to jest takiego, który zachowuje się w sposób niezgodny z oczekiwaniami badacza i ankiera. W literaturze przedmiotu nie spotkałem się z rozważaniami dotyczącymi takiej problematyki.

## METODOLOGIA

Badanie zostało przeprowadzone w okresie od listopada 2009 do sierpnia 2010 r. Objęto nim 12 polskich komercyjnych ośrodków badawczych realizujących badania CATI<sup>7</sup>: *4P research mix*, *ASM – Centrum Badań i Analiz Rynku*, *ARC Rynek i Opinia*, *Expert-Monitor* (obecnie *Kantar Media Intelligence*), *GfK Polonia*, *IMAS International*, *IPSOS*, *Grupa IQS*, *MillwardBrown SMG/KRC*, *PBS DGA*, *Pentor Research International* (obecnie *TNS Polska*) i *TNS OBOP* (obecnie *TNS Polska*). Kryterium doboru agencji był fakt posiadania certyfikatu jakości w kategorii wywiadu CATI, przyznanego przez Organizację Firm Badania Opinii i Rynku (OFBOR) w wyniku VII edycji (2008 r.) inspekcji stosowania standardów Programu Kontroli Jakości Pracy Ankierów (PKJPA). Certyfikatem takim legitymowało się 18 firm – w sześciu agencjach realizacja badania – z uwagi na odmowę lub likwidację studia CATI – nie powiodła się.

W każdej z instytucji przy użyciu ankiety (papierowej lub elektronicznej<sup>8</sup>) przebadalem możliwie liczną grupę ankierów CATI, docierając w sumie do 846 osób. Odsetki realizacji w poszczególnych firmach były mocno różnicowane i wahały się od 20 do 100%. Za 100% przyjęto liczbę ankierów, którzy w przeciągu miesiąca poprzedzającego początek badania w danej firmie przynajmniej raz byli w pracy.

<sup>7</sup> Takie samo badanie przeprowadziłem też w dwóch ośrodkach zagranicznych: *Capacent Gallup* (w Islandii) oraz urzędzie statystycznym Norwegii: *Statistik sentralbyra (Statistics Norway)*. Realizacja zagranicznej części projektu została sfinansowana z dwóch grantów otrzymanych przeze mnie ze środków pomocy Mechanizmu Norweskiego EOG i Norweskiego Mechanizmu Finansowego w ramach Funduszu Stypendialnego i Szkoleniowego. W każdej instytucji odbyłem 10-dniową wizytę studyjną. Wyniki modułu zagranicznego nie są prezentowane w tym artykule.

<sup>8</sup> W większości agencji narzędzie badawcze miało postać papierową; ankierzy wypełniali kwestionariusze podczas pracy (w specjalnie do tego przeznaczonym czasie lub też w przerwach między wykonywanymi połączeniami) i następnie oddawali mi je. Dwie agencje nie zgodziły się na taką formę badania – wykorzystano tam narzędzie elektroniczne: ankierzy wypełniali ankietę zamieszczoną w Internecie lub też w wewnętrznym systemie informatycznym firmy.

Uzupełnieniem części ilościowej projektu były indywidualne wywiady jakościowe realizowane z ankieterami CATI. Do wywiadów tych rekrutowałem ankieterów pracujących w zawodzie co najmniej jeden rok, dbając by reprezentowani byli współpracownicy wszystkich agencji, które uczestniczyły w części ilościowej oraz by zapewnić zróżnicowanie badanych pod względem wieku, płci i profilu wykształcenia. Rekrutacja do wywiadów odbywała się na podstawie informacji zostawionych przez badanych w kwestionariuszu wykorzystanym w części ilościowej – ankieterzy byli tam informowani o planowanych wywiadach i proszeni o zostawienie swoich danych kontaktowych (numeru telefonu lub adresu mailowego). Z niektórymi osobami, które takie informacje zamieściły, nawiązywałem później indywidualnie kontakt i umawiałem się na wywiad. Zrealizowałem 32 wywiady jakościowe.

Poniżej analizuję odpowiedzi na jedno z pytań ankiety, dotyczące stresujących czynności ankieterskich. W dalszej części prezentuję dane z wywiadów jakościowych opisujące strategie postępowania wdrażane przez ankieterów w kontakcie z trudnymi respondentami<sup>9</sup>. Wyniki badania – co oczywiste – nie są reprezentatywne dla całej zbiorowości ankieterów realizujących w Polsce sondaże telefoniczne. Oprócz firm posiadających certyfikat PKJPA funkcjonują agencje, które takiego znaku jakości nie mają. Co więcej, wiele instytutów badawczych zleca obecnie realizację terenową swoich projektów firmom zewnętrznym (wyspecjalizowanym w działaniach realizacyjnych), rezygnując z posiadania własnych sieci ankieterskich. Taki *outsourcing* ma miejsce w szczególności w obszarze badań telefonicznych, prawdopodobnie z uwagi na duże koszty, jakie pochłania budowa i utrzymanie infrastruktury CATI.

## CZYNNOŚCI STRESUJĄCE

W jednym z pytań kwestionariusza przedstawiłem listę 11, potencjalnie kłopotliwych, czynności ankieterskich. W odniesieniu do każdej z nich ankieter proszony był o ocenę, w jakim stopniu jest ona dla niego stresująca – czy jest ona: w ogóle niestresująca, raczej niestresująca, raczej stresująca, czy też bardzo stresująca. Ankieter mógł też zaznaczyć opcję „trudno powiedzieć” lub „nie było takiej sytuacji”.

<sup>9</sup> Kwestionariusz ankiety zawierał również pytania związane ze szkoleniem wprowadzającym oraz przeszłością zawodową badanych, mierzył też poziom satysfakcji z wykonywanej pracy. Z kolei scenariusz wywiadu przewidywał również rozmowę na temat postrzeganego wpływu nieosobistego kontaktu z respondentem oraz różnych cech wywiadu CATI (np. długość, rodzaje zadawanych pytań, tematyka, itp.) na prawdopodobieństwo pojawienia się trudności w kontakcie z respondentem.

Chcąc zidentyfikować wymiary czynności stresujących przeprowadziłem eksploracyjną analizę czynnikową metodą głównych składowych. Do określenia liczby czynników użyłem kryterium wykresu osypiska [Bedyńska, Cypryańska 2007], na podstawie którego ustaliłem, że należy wyodrębnić dwa czynniki. Po wykonaniu analizy z rotacją ortogonalną metodą Varimax stwierdziłem, że pierwszy czynnik tworzą cztery itemy odnoszące się do czynności związanych z ryzykiem przerwania wywiadu przez respondenta, natomiast drugi czynnik silnie ładuje cztery itemy opisujące czynności prowadzące do obciążenia poznawczego ankietera. Pozostałe trzy itemy usunąłem z analizy, gdyż nie były one silnie nasycone żadnym z wyodrębnionych czynników<sup>10</sup>. Szczegółowe wyniki prezentuję w tabeli 1.

TABELA 1. Stresujące czynności ankierskie – analiza czynnikowa

| Czynnik / czynność ankierska  | Średnia arytmetyczna (skala od 1: w ogóle nie stresujące do 4: bardzo stresujące) <sup>1</sup> |
|---|--|
| 1   | 2  |
| Czynnik 1: czynności związane z ryzykiem przerwania wywiadu przez respondenta   | 2,64   |
| Przeprowadzanie wywiadu z respondentem „doczepiającym się” do wielu pytań, narzekającym na ich bezsensowność, niepoprawność, itp. (0,807 <sup>2</sup> ) | 2,85   |
| Aranżowanie wywiadu z respondentem agresywnym – krzyżującym, odgrażającym się, itp. (0,791)   | 2,85   |
| Przeprowadzanie długiego wywiadu, trwającego ponad 30 minut (0,586)   | 2,62   |
| Konwertowanie odmowy – nakłanianie do udziału w badaniu respondenta, który w czasie poprzedniej rozmowy odmówił (łagodna odmowa) (0,710)                | 2,26   |
| Czynnik 2: czynności powodujące obciążenie poznawcze ankietera  | 2,03   |
| Przeprowadzanie wywiadu z respondentem słabo słyszającym (0,773)  | 2,54   |
| Przeprowadzanie wywiadu z respondentem z poważną wadą wymowy – jękającym się, mówiącym niewyraźnie, itp. (0,809)  | 2,26   |
| Notowanie rozbudowanej odpowiedzi respondenta na pytanie otwarte (0,627)  | 1,82   |

<sup>10</sup> Usuniętymi itemami były: „Dopytywanie respondenta odpowiadającego nie na temat, niejednoznacznie, lakonicznie, itp.”; „Aranżowanie wywiadu z respondentem twierdzącym, iż jego numer telefonu jest zastrzeżony i ośrodek nie ma prawa do niego dzwonić (próba numerów generowanych losowo)”; „Aranżowanie wywiadu z respondentem podejrzewającym, że rzeczywistym celem rozmowy jest nie przeprowadzenie wywiadu, a nakłonienie do zakupu jakiejś rzeczy”.

| 1   | 2    |
|---|------|
| Realizowanie wielu różnych projektów w trakcie jednej sesji (0,475) | 1,53 |

Wyznacznik = 0,027; K-M-O = 0,854<sup>11</sup>; rotacja Varimax

Czynnościami bardziej stresującymi dla ankietera okazały się te związane z ryzykiem przerwania wywiadu przez respondenta. Ryzyko to spowodowane jest cechami charakterologicznymi respondenta (skłonność do agresji, wytykania nieprawidłowości w narzędziu badawczym, niechęć do uczestniczenia w badaniu) lub też cechą samego wywiadu – jego długością. Czynności mniej stresujące dla ankieterów to te, które powodują ich obciążenie poznawcze. Zaliczyć możemy do nich przede wszystkim czynności odnoszące się do zakłóceń komunikacyjnych na linii ankieter-respondent oraz koordynacji słuchania i pisania. Kłopotliwe – choć w mniejszym stopniu – jest też dla ankieterów częste przełączanie się między realizowanymi projektami.

Chcąc dokładniej zanalizować wyniki, szukałem zależności między stopniem zestresowania ankietera danymi czynnościami a cechami społeczno-demograficznymi i cechami opisującymi charakter pracy w studiu. Wyniki analiz prezentuję w tabelach 2–5<sup>12</sup>.

TABELA 2. Poziom zestresowania a płeć ankietera

| Wymiary stresu   | Płeć ankietera |           |
|--|----------------|-----------|
|  | kobieta        | mężczyzna |
| Czynności związane z ryzykiem przerwania wywiadu przez respondenta | 2,73           | 2,47      |
| Czynności powodujące poznawcze obciążenie ankietera                | 2,04           | 2,01      |

TABELA 3. Poziom zestresowania a wiek ankietera

| Wymiary stresu   | Wiek ankietera (lata życia) |       |       |             |
|--|-----------------------------|-------|-------|-------------|
|  | do 22                       | 23–24 | 25–28 | 29 i więcej |
| Czynności związane z ryzykiem przerwania wywiadu przez respondenta | 2,56                        | 2,53  | 2,70  | 2,78        |
| Czynności powodujące poznawcze obciążenie ankietera                | 2,06                        | 1,96  | 1,97  | 2,12        |

<sup>11</sup> Wyznacznik i K-M-O (miara adekwatności doboru próby Kaisera, Mayera i Olkina) to miary testujące właściwości macierzy korelacji [Bedyńska, Cypryńska 2007].

<sup>12</sup> W tabelach podano wartości średnich arytmetycznych (skala od 1: w ogóle nie stresujące do 4: bardzo stresujące; z analiz wykluczono odpowiedzi „trudno powiedzieć” i „nigdy nie było takiej sytuacji”).

TABELA 4. Poziom zestresowania a doświadczenie ankierskie

| Wymiary stresu   | Staż pracy ankietera w CATI (miesiące) |      |       |             |
|--|--|------|-------|-------------|
|  | do 4                                   | 5–11 | 12–23 | 24 i więcej |
| Czynności związane z ryzykiem przerwania wywiadu przez respondenta | 2,51                                   | 2,65 | 2,68  | 2,70        |
| Czynności powodujące poznawcze obciążenie ankietera                | 2,05                                   | 2,00 | 2,01  | 2,02        |

TABELA 5. Poziom zestresowania a obciążenie pracą

| Wymiary stresu   | Obciążenie pracą ankietera <sup>3</sup> |         |      |
|--|---|---------|------|
|  | małe                                    | średnie | duże |
| Czynności związane z ryzykiem przerwania wywiadu przez respondenta | 2,71                                    | 2,64    | 2,58 |
| Czynności powodujące poznawcze obciążenie ankietera                | 2,02                                    | 2,05    | 1,90 |

Większe różnicowanie poziomu zestresowania wśród różnych kategorii ankierów widoczne jest w przypadku czynnika 1, a więc tego odnoszącego się do czynności mogących skutkować przerwaniem wywiadu przez respondenta. Jak widać, bardziej wrażliwe na tego typu sytuacje są kobiety, osoby starsze, ankierzy o dłuższym stażu pracy oraz o mniejszym obciążeniu pracą. Niewykluczone, iż ankierzy starsi i dłużej pracujący w studiu CATI, traktują to zajęcie bardziej poważnie i stanowi ono główne źródło ich utrzymania. W konsekwencji osoby te w większym stopniu obawiają się zdarzeń, które mogą być traktowane w kategorii porażki. Z kolei mniejszy poziom zestresowania wśród ankierów bardziej obciążonych pracą wyjaśnić można jako efekt „wyrobienia się”, pewnego unieważnienia na czynniki potencjalnie stresogenne. Choć może być też tak, że mniejsza podatność na stres jest u tych osób cechą osobowościową, a wykonywanie pracy z większą intensywnością to niejako rezultat posiadania takiej konstrukcji psychicznej.

Analizowane cechy w mniejszym stopniu różnicują poziom zestresowania w przypadku czynnika związanego z czynnościami mogącymi powodować obciążenie poznawcze ankietera. Najistotniejszymi zmiennymi wydają się tu być wiek i intensywność pracy. Osoby zakwalifikowane do brzegowych kohort wiekowych (do 22 lat i od 29 lat) wykazują większy stopień zestresowania aniżeli osoby w wieku od 23 do 28 lat. Ankierzy najbardziej obciążeni pracą są mniej zestresowani w porównaniu do ankierów pracujących z mniejszą intensywnością.

## TRUDNI RESPONDENCI

Przejdę teraz do opisu wyników części jakościowej badania. Jak już wspominałem, interesowało mnie, jakie środki zaradcze wdrażają ankieterzy w sytuacji kontaktu z respondentami niewspółpracującymi. Wyróżniłem następujące kategorie trudnych respondentów:

- 1) respondenci niechętni – odmawiający uczestnictwa w wywiadzie;
- 2) respondenci zirytowani, narzekający na wywiad;
- 3) respondenci mający problem ze zrozumieniem treści pytań;
- 4) respondenci cechujący się dużą aktywnością uboczną (komentujący pytania, uciekający w dygresje, itp.).

(1) Większość badanych ankieterów przyznaje, że w przypadku natrafienia na respondenta odmawiającego udziału w wywiadzie, kalkuluje „opłacalność” nakłaniania rozmówcy do współpracy. Jeżeli należy on do kategorii respondentów łatwodostępnych (czyli penetracja w populacji osób o wskazanych cechach jest duża) i/lub baza numerów, jakie są w próbie jest liczna, ankieterzy na ogół rezygnują z przekonywania, żegnają się z rozmówcą i dzwonią pod inny numer. Jeśli jednak baza numerów będących w dyspozycji studia jest ograniczona i/lub do badania rekrutowani są ludzie o cechach mało popularnych, większość ankieterów podejmuje próbę nakłonienia rozmówcy do współpracy.

*Jeżeli jest to zwykłe badanie masowe, to podchodzę do tego na zasadzie „nie ten, to następny”, tutaj mniej szanuje się bazę respondentów. Jeżeli on się zgodził w pierwszych słowach, to rozmowa z nim jest stratą czasu. Jeżeli jest to badanie, gdzie baza jest okrojona, mam powiedzieć 10 kontaktów i nie ma określonej kwoty, mam zrobić jak najwięcej się da, to jest to o wiele trudniejsze, muszę się tutaj bardziej postarać.* (kobieta, 23 lata, ankieterka CATI od dwóch lat)

*Generalnie nie przesadzam z namawianiem. Zawsze się boję tego, że zacznę wywiad, a respondent mi odmówi w połowie – to jest wtedy stracony czas. Po co ja mam się z nim męczyć i ryzykować, że mi odłoży słuchawkę, jak może mi się zaraz jakiś inny respondent bez problemu zgodzić.* (kobieta, 25 lat, ankieterka CATI od półtora roku)

Niekiedy podejmowanie prób nakłonienia respondenta wynika też z nastroju ankietera (ma on ochotę podyskutować z kimś, posprzeczać się, sprawdzić, na ile perswazyjna jest jego argumentacja) lub też z chęci umilenia sobie czasu (rozmowa z respondentem jest przyjemniejsza niż ciągle przedstawianie się i wygłaszanie formuły aranżacyjnej). W takich sytuacjach jednak zachowanie ankieterów zdaje się być zależne od zdecydowania, z jaką respondent wstępnie omawia – jeżeli odmowa jest „twarda”, większość ankieterów kończy taki kontakt

– oraz od oceny tego, jak sprawnie respondent poradziłby sobie z wywiadem – jeżeli ankieterzy spodziewają się trudności w odpowiadaniu na pytania, raczej powstrzymują się od nakłaniania.

*Jak przychodzę do pracy na 8 godzin, to przez pierwsze 4 nakłaniam, bo lubię to robić, poza tym czas szybciej biegnie, jak rozmawiamy z osobą, a nie słuchamy sygnału [...]. Później, jak jestem zmęczona, to przy odmowie mówię „dziękuję bardzo”.* (kobieta, 24 lata, ankieterka CATI od roku)

*Też wiele zależy od mojego humoru. Jak mam dobry humor, to się lubię z takimi uparciuchami podroczyć i niekiedy udaje mi się ich namówić [...]. A jak jestem w gorszym nastroju, to mi się nie chce i odpuszczam.* (mężczyzna, 25 lat, ankieter CATI od roku)

*Dwa przypadki, kiedy wiem z góry, że nic nie ugram i odpuszczam: silny, zdecydowany głos mówiący „nie” i stara babuleńka, która nic nie rozumie i nawet jak się zgodzi, to nie odpowie na żadne pytanie.* (mężczyzna, 25 lat, ankieter CATI od dwóch lat)

Generalnie można stwierdzić, że perswazyjność aranżacji w wywiadach telefonicznych jest stosunkowo słaba. Dominuje dążenie do realizacji jak największej liczby wywiadów w jak najkrótszym czasie. Jak przyznają niektórzy ankieterzy, spora część połączeń kończących się odmową mogłaby skutkować realizacją wywiadu, gdyby wysiłek aranżacyjny podejmowany przez dzwoniących był większy.

*Praktyka mi pokazała, że za parę godzin respondenci, którzy właśnie odmówili, mogą się zgodzić i jeszcze okazać się idealnymi respondentami. Po prostu czasem dzwoniemy w beznadziejnych sytuacjach i stąd takie reakcje. Dlatego ja często takie przypadki wrzucam na za 3 godziny i wpisuję „niemiły facet”. Potem się okazuje, że w tym czasie, kiedy dzwoniłem, opieprzał go szef, a teraz człowiek siedzi w domu, je obiad i mu się nudzi. [...] Kiedyś mieliśmy bardzo ograniczoną bazę numerów, która wykończyła nam się po jednym dniu. Szefostwo zrobiło tak, że na drugi dzień wrzucili znowu te same numery, mówiąc, że baza jest nowa. I nagle z tych odrzuconych numerów, 1/3 się zgodziła.* (mężczyzna, 25 lat, ankieter CATI od roku)

(2) Osoby poddenerwowane, narzekające na wywiad to kolejna kategoria respondentów przysparzająca ankieterom kłopotów. Najczęściej zdenerwowanie to wynika z długości rozmowy, tzn. w opinii respondentów wywiad, w którym uczestniczą powinien być krótszy. Zdecydowana większość ankieterów „zabezpiecza się” przed taką sytuacją, mówiąc w trakcie aranżacji, jak długo będzie trwała rozmowa. Kiedy respondent reaguje zdenerwowaniem, ankieterzy informują, ile minut minęło od początku wywiadu i że rozmówca zgodził się na określony czas. Strategia taka – jak wynika z relacji ankieterskich – powinna być jednak stosowana z rozwagą:



*Mówi się w aranżacji, że wywiad trwa 10 minut. Ale to jest czas prawdziwy w przypadku respondenta, który odpowiada sprawnie. Jak ktoś się zastanawia, taki wywiad przedłuża się do 20 minut. To sprawia, że niektórzy rzucają słuchawkę. [...] Zdarzyła mi się raz taka sytuacja, że respondent odmierzał mi czas. Powiedziałam, że wywiad będzie trwał 10 minut. On po 10 minutach powiedział: „przeznaczyłem dla pani tyle czasu, ile pani chciała; dziękuję bardzo, do widzenia”.* (kobieta, 56 lat, ankieterka CATI od półtora roku)

Ankieterzy przyznają, że zdenerwowanie respondenta w wielu przypadkach zakończyć się może przerwaniem połączenia. Wyczuwając w głosie respondenta narastającą irytację, ankieterzy starają się „zmiękczyć” rozmówcę, podkreślając, że zostało już niewiele do końca lub też zwracając uwagę, że jedynie całkowicie zrealizowany wywiad może być uznany – w razie przerwania rozmowy, czas przeznaczony do tej pory przez badanego byłby czasem straconym. Dość często stosowaną tutaj strategią jest też zwiększanie tempa mówienia oraz – jeżeli rezygnacja respondenta wydaje się nieunikniona – wyjście z propozycją rozłożenia wywiadu na raty i dokończenia rozmowy w innym terminie.

W sytuacji, gdy respondent przerywa rozmowę w trakcie wywiadu i rozłącza się, ankieterzy albo postępują zgodnie z procedurą, opisując w systemie zdarzenie lub też oddzwaniają do respondenta udając, że przerwanie rozmowy było skutkiem usterki technicznej, a nie intencjonalnym działaniem respondenta. Co zaskakujące, taki zabieg jest dość często stosowany i – w opinii ankieterów – cechuje go spora skuteczność.

*Jeżeli respondent rzuca mi słuchawkę, to ja robię tak, że za chwilę do niego oddzwaniam i mówię, że coś przerwało i powtarzam aranżację. Wtedy albo się zgadza, albo nie, ale z reguły jest mu głupio i się zgadza.* (kobieta, 25 lat, ankieterka CATI od półtora roku)

*Jak ktoś mi rzuci słuchawkę, to ja dzwonię jeszcze raz i się pytam: „pani rzuciła słuchawkę, czy coś nas rozłączyło?” Czasami to skutkuje, ale jeżeli respondent rzuci słuchawkę drugi raz, to jest spuszczone, nie ma co go umawiać i utrudniać pracy innym ankieterom.* (kobieta, 23 lata, ankieterka CATI od dwóch lat)

Niektórzy ankieterzy zwracają uwagę, że zdenerwowanie respondenta może być też pochodną braku asertywności. Respondenci tacy nie umieją powiedzieć wprost, że nie chcą brać udziału w badaniu, próbują się „wykrecać”, zwodzą ankietera, ale wreszcie – pod naciskiem – wyrażają zgodę na wywiad. Później na pytania ankietera reagują irytacją, niekiedy w furii przerywając rozmowę.

*Ludzie bardzo często kłamią. Godzina 21, na zewnątrz 20 stopni mrozu. Starsza pani mówi, że wychodzi. Ona nie wychodzi, tylko nie ma odwagi powiedzieć, że nie chce*

*brać udziału w wywiadzie. Ważna jest asertywność.* (kobieta, 50 lat, ankieterka CATI od czterech lat)

*Główna odpowiedź ludzi, którzy nie chcą wziąć udziału w ankiecie, to jest, że wychodzą. Nawet jak się na komórkę dzwoni, to oni wychodzą. Ja się wtedy pytam, kiedy wrócę.* (mężczyzna, 52 lata, ankieter CATI od trzech lat)

(3) Ankieterzy zgodnie przyznają, że bardzo często w swojej pracy natrafiają na respondentów mających trudności ze zrozumieniem pytań. Strategią uruchamianą czasami w takiej sytuacji jest ściśle trzymanie się zaleceń (powtarzanie pytania wolniejszym tempem, ewentualne zaznaczenie kategorii „trudno powiedzieć”, w skrajnych przypadkach podziękowanie respondentowi i przerwanie wywiadu).

*Czasami są osoby mocno niekomunikatywne, najczęściej wynika to z wieku, ale nie zawsze. One po prostu nie rozumieją, co się do nich mówi. Taka rozmowa nie ma najmniejszego sensu, bo musiałabym wszystkie pytania zadawać po swojemu. Jeżeli po kilku pierwszych ekranach orientuję się, że ktoś w ogóle nie ma pojęcia, o czym ja mówię – rezygnuję, sama kończę rozmowę.* (kobieta, 61 lat, ankieterka CATI od dwóch i pół lat)

Zdarzają się jednak (i takie odpowiedzi przeważają) próby mniejszej lub większej – destandardyzacji procedury realizacyjnej. Polega ona na ogół na reformułowywaniu pytań tak, aby były one bardziej przystępne dla respondenta i wyjaśnianiu niezrozumiałych terminów. Rodzaj wdrażanej strategii wydaje się być zależeć od firmy, w jakiej pracuje ankieter. Choć wszyscy są świadomi znaczenia ujednolicenia sposobu przeprowadzania wywiadu, w większości ośrodków panuje przyzwolenie na odchodzenie od zaleceń w przypadku wystąpienia trudności w zrozumieniu pytań – przyzwolenie takie ma zwykle charakter milczący, choć w pojedynczych przypadkach sami superwizorzy instruuja ankieterów, aby reformułowywać pytania. Jedynie w 2 ośrodkach – jak wynika z relacji ankieterów – pilnuje się tego, aby zachowanie ankieterów w omawianych sytuacjach było idealnie podręcznikowe.

*Bardzo często jest tak, że pytania w ankiecie musimy upraszczać [...], jak ludzie nie rozumieją pytań, to się denerwują, chcą się rozłączyć. To oczywiście nie jest zalecenie oficjalne, ale nieoficjalne owszem. Nikt się do tego nie doczepia. Jeśli rozmawiam z tłu-kiem, osobą, która nie rozumie większości słów, to mam dwie możliwości – zakończyć wywiad, bo respondent nie jest w stanie odpowiadać na pytania, lub dostosować się do respondenta.* (mężczyzna, 24 lata, ankieter CATI od roku)

Warto zauważyć, że w opinii prawie wszystkich rozmówców pytania w skryptach CATI często są sformułowane przy użyciu nadmiernie skomplikowanego

słownictwa i składni – są przeintelektualizowane, niedostosowane do możliwości intelektualnych przeciętnego respondenta. Gdyby – jak uważają ankieterzy – przywiązywano większą wagę to tego, jak skonstruowane jest narzędzie badawcze (odsłuchiowano wywiadów, rozmawiano z ludźmi pracującymi w studiu), do przypadków odchodzenia od reguł standaryzacji by nie dochodziło.

*Czasami pytania są tak pokracznie sformułowane, że nawet ja nie wiem, o co chodzi. Więc jak ci biedni respondenci z podstawowym czy zasadniczym wykształceniem mają je zrozumieć? [...] Żaden badacz, żaden superwizor, który nie siedział na słuchawkach, nie wie tego, co my – ankieterzy – wiemy. (mężczyzna, 62 lata, ankieter CATI od trzech lat)*

*Zdarzają się czasami tak długie pytania, że jak skończę je czytać, to sama nie pamiętam, co było na początku. Respondent – nie widząc tekstu – tym bardziej. (kobieta, 69 lat, ankieterka CATI od dwóch lat)*

(4) Ostatni analizowany typ trudnego respondenta to rozmówca cechujący się dużą aktywnością uboczną – komentujący pytania, uzasadniający swoje odpowiedzi, uciekający w dygresje, itp. Jak przyznają ankieterzy, obcowanie z takimi respondentami jest mocno kłopotliwe, gdyż z jednej strony obowiązuje reguła uprzejmości wobec badanego, z drugiej nadzorujący pracę w studiu CATI zwykle pilnują czasu, jaki poświęcany jest na realizację pojedynczego przypadku – zbyt długi wywiad generuje koszty. W konsekwencji ankieterzy próbują stosować się jednocześnie do tych dwóch wymogów, balansując między byciem uprzejmym i byciem efektywnym. W skrajnych przypadkach – podobnie jak w przypadku respondentów z trudnościami poznawczymi – wywiady są przerywane przez samych ankieterów.

*Dygresje to koszmar każdego ankietera. Superwizor mówi, że ankieta ma trwać 5 minut, a ona trwa 10 – dlatego, że respondent zaczyna opowiadać. Ja mam taką metodę, że czekam do pierwszego wdechu i wtedy wchodzę z następnym pytaniem. Niekiedy działa, a niekiedy ludzie są tak niereformowalni, że 5-minutowa ankieta trwa pół godziny. (mężczyzna, 24 lata, ankieter CATI od roku)*

*Jak respondent się rozwodzi i opowiada o różnych sprawach, to ja piluję paznokcie, książkę czytam, albo włączam „mute” i gadam z koleżanką. Z reguły dają się wygadać, ale jak trwa to strasznie długo, to mówię respondentowi, że nie chcę zajmować dłużej niż potrzeba jego czasu i proponuję, abyśmy wrócili do pytań. (kobieta, 25 lat, ankieterka CATI od półtora roku)*

*Ścinanie jest związane z ryzykiem – może skończyć się odmową. Trochę trzeba wysłuchać, pozwolić się wygadać temu respondentowi. Niekiedy jest jednak tak, że przedmiot badania wywołuje u respondenta negatywne emocje i zaczyna on bluźnić. Wtedy*

*ucinam, nie jestem zainteresowana tym, żeby taki rynsztok na mnie sphywał.* (kobieta, 61 lat, ankieterka CATI od dwóch i pół lat)

*Zdarza się, że trzeba przerwać taki wywiad. Ktoś mówi „niech mnie pani teraz słucha, ja chcę się wypowiedzieć, nie będę na te głupie pytania odpowiadał”. Jeżeli ktoś nie będzie odpowiadał na głupie pytania, to ja rozumiem, że on nie jest zainteresowany wywiadem – i taką rozmowę kończę.* (kobieta, 25 lat, ankieterka CATI od trzech i pół lat)

Zachowaniem praktykowanym przez wielu ankieterów jest też proponowanie „rozgadanemu” respondentowi, aby wszystkie swoje uwagi zachował do końca wywiadu – wówczas zostaną one zarejestrowane.

*Zdarza się, że ktoś na początku deklaruje, że ma bardzo mało czasu, a później się rozgaduje i opowiada o różnych sprawach przy każdym pytaniu. Wtedy mówię, że pod koniec wywiadu będę miała możliwość zapisania swobodnych opinii, a teraz proszę jedynie o udzielenie odpowiedzi na skali. Na końcu już nikt o tym nie pamięta. Zresztą ja na to liczę, szczerze mówiąc. Niekiedy w skrypcie jest takie pytanie „czy coś jeszcze chciałby pan dodać? Stąd to wzięłam.* (kobieta, 61 lat, ankieterka CATI od dwóch i pół lat)

## ZAKOŃCZENIE

Wyniki badania wskazują na istnienie dwóch grup sytuacji wyzwalających reakcje stresowe u ankieterów CATI. Do pierwszej z nich zaliczyć możemy czynności ankierskie związane z ryzykiem przerwania wywiadu przez respondenta (przeprowadzanie wywiadów długich, kontakt z respondentami agresywnymi oraz kwestionującymi poprawność zadawanych pytań), do drugiej – czynności powodujące znaczne poznawcze obciążenie ankietera (przede wszystkim przeprowadzanie wywiadów z respondentami mającymi problem ze słuchem i niewyraźnie mówiącymi).

Przytoczone wyniki odsłaniają również popularne strategie uruchamiane przez ankieterów w sytuacji natrafienia na trudnego respondenta. Najbardziej zastanawiające jest – jak sądzę – postępowanie ankieterów właściwe dla wywiadów prowadzonych z osobami mającymi trudność w zrozumieniu treści zadawanych pytań. Choć destandaryzowanie procedury realizacyjnej, do którego przyznają się ankieterzy, jest prawdopodobnie o wiele słabsze i rzadziej stosowane aniżeli w wywiadach terenowych – *face-to-face* – jest ono źródłem błędów i obniża wartość zebranego materiału.

Moim zdaniem warto zwrócić uwagę na fakt, iż – jak wynika z wypowiedzi ankieterów – w zdecydowanej większości firm badawczych osoby nadzorujące pracę studia CATI przyzwalają na stosowanie zabiegów destandaryzacyjnych. Niewykluczone, że superwizorzy podzielają wyrażane powszechnie przez an-

kieterów opinie o nieprzystawalności języka pytań oraz zadań umieszczanych w skryptach CATI do możliwości intelektualnych przeciętnego respondenta i – rozumiejąc trudną sytuację ankieterów – nie egzekwują wymogu ścisłej standaryzacji. Być może właściwe byłoby uwzględnienie sugestii ankieterów i zwrócenie większej uwagi na kwestię przyjaznego dla respondenta redagowania pytań w skrypcie, jak również wyposażenie ankieterów w instrukcje regulujące (i zarazem standaryzujące) sposób zachowania w przypadku kontaktu z kłopotliwym rozmówcą.

## BIBLIOGRAFIA

- Bedyńska S., Cypriańska M. [2007], *Zaawansowane sposoby tworzenia wskaźników: zastosowanie analizy czynnikowej oraz analizy rzetelności pozycji*, [w:] Bedyńska S., Brzezicka A. (red.), *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii*, Warszawa: SWPS Academica.
- Biemer P.P., Lyberg L.E. [2003], *Introduction to survey quality*, Hoboken: John Wiley & Sons.
- Biuletyn informacyjny Centralnego Biura Spisowego* [2011], vol. 48, Warszawa: GUS.
- Chojnowski P., Wódkowski A. [2011], *Rynek badań marketingowych z perspektywy 2011 roku*, [w:] Katalog PTBRiO 2011/12, edycja XVI, Warszawa: PTBRiO.
- Conrad F.G., Schober M.F. [2000], *Clarifying question meaning in a household telephone survey*, "Public Opinion Quarterly", vol. 64 (1).
- Couper M.P., Groves R.M. [2002], *Introductory interactions in telephone surveys and non-response*, [w:] Maynard D.W., Houtkoop-Steenstra H., Schaeffer N.C., van der Zouwen J. (red.), *Standardization and tacit knowledge. Interaction and practice in the survey interview*, New York, John Wiley & Sons.
- de Leeuw E.D. [2002], *Data quality in mail, telephone and face to face surveys*, Amsterdam: TT-Publikaties.
- Fazzi G., Martire F., Pitrone M.C. [2009], *Hanging by a thread: the telephone interviewers tell (their) story*, JSM Proceedings (Survey Research Methods Section), Alexandria: American Statistical Association.
- Fowler F.J., Mangione T.W. [1990], *Standardized survey interviewing: minimizing interviewer-related error*, Newbury Park: Sage.
- Frey J.H. [1989], *Survey research by telephone*, Newbury Park: Sage.
- Groves R.M. [1979], *Actors and questions in telephone and personal interview surveys*, "Public Opinion Quarterly", vol. 43 (2).
- Groves R.M. [1990], *Theories and methods of telephone surveys*, "Annual Review of Sociology", vol. 16.
- Groves R.M., Fowler F.J., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. [2009], *Survey methodology*, Hoboken: John Wiley & Sons.
- Groves R.M., Magilavy L.J. [1986], *Measuring and explaining interviewer effects in centralized telephone surveys*, "Public Opinion Quarterly", vol. 50 (2).

- Groves R.M., O'Hare B.C., Gould-Smith D., Benki J., Maher P. [2008], *Telephone interviewer voice characteristics and the survey participation decision*, [w:] Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japac L., Lavrakas P.J., Link M.W., Sangster R.L. (red.), *Advances in telephone survey methodology*, Hoboken: John Wiley & Sons.
- Herzog A.R., Rodgers W.L. [1988], *Interviewing older adults: mode comparison using data from a face-to-face survey and a telephone resurvey*, "Public Opinion Quarterly", vol. 52(1).
- Holbrook A.L., Green M.C., Krosnick J.A. [2003], *Telephone versus face-to-face interviewing of national probability samples with long questionnaires. Comparisons of respondent satisficing and social desirability response bias*, "Public Opinion Quarterly", vol. 67 (1).
- Hołodakowski W. [2006], *Nie(d)ocenione CATI*, "One Way Mirror", vol. 24.
- Hyman H.H. [1954], *Interviewing in social research*, Chicago: University of Chicago Press.
- Jordan L.A., Marcus A.C., Reeder L.G. [1980], *Response styles in telephone and household interviewing: a field experiment*, "Public Opinion Quarterly", vol. 44 (2).
- Kalsbeek W.D., Agans, R.P. [2007], *Sampling and weighting in household telephone surveys*, [w:] Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japac L., Lavrakas P.J., Link M.W., Sangster R.L. (red.), *Advances in telephone survey methodology*, Hoboken: John Wiley & Sons.
- Keeter S., Kennedy C., Clark A., Tompson T., Mokrzycki M. [2007], *What's missing from national landline RDD surveys? The impact of the growing cell-only population*, "Public Opinion Quarterly", vol. 71 (5).
- Kelly J., Link M.W., Petty J., Hobson K., Cagney P. [2008], *Establishing a new survey research call center*, [w:] Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japac L., Lavrakas P.J., Link M.W., Sangster R.L. (red.), *Advances in telephone survey methodology*, Hoboken: John Wiley & Sons.
- Kormendi E. [1988], *The quality of income information in telephone and face to face surveys*, [w:] Groves R.M., Biemer P.P., Lyberg L.E., Massey J.T., Nicholls II W.L., Waksberg J. (eds), *Telephone survey methodology*, New York, John Wiley & Sons.
- Krosnick J. [1999], *Survey research*, "Annual Review of Psychology", vol. 50.
- Kubiak A., Przybyłowska I., Rostocki W.A. [1992], *Spoleczna przestrzeń wywiadu kwestionariuszowego*, „Analizy i próby technik badawczych w socjologii”, vol. 9.
- Kujawski M., Mirowski S., Hołodakowski W. [2008], *Badania CATI wobec spadku penetracji telefonów stacjonarnych*, "One Way Mirror", vol. 28.
- Lavrakas P.J. [1993], *Telephone survey methods: sampling, selection, and supervision*, Thousand Oaks: Sage.
- Loosveldt G. [1997], *Interaction characteristics of the difficult-to-interview respondent*, "International Journal of Public Opinion Research", vol. 9.
- Oksenberg L., Cannell C. [1988], *Effects of interviewer vocal characteristics on nonresponse*, [w:] Groves R.M., Biemer P.P., Lyberg L.E., Massey J.T., Nicholls II W.L., Waksberg J. (red.), *Telephone survey methodology*, New York: John Wiley & Sons.
- Polskie standardy jakości realizacji badań rynku i opinii społecznej w terenie* [2010], Warszawa: OFBOR.
- Sawiński Z. [1996], *Sondaże telefoniczne*, „ASK. Społeczeństwo, Badania, Metody”, vol. 1/1996..
- Sawiński Z. [2005], *Wywiady osobiste z komputerem przenośnym (CAPI)*, [w:] Sztabiński P.B., Sawiński Z., Sztabiński F. (red.), *Fieldwork jest sztuką*, Warszawa, IFiS PAN.

- Steeh Ch. [2008], *Telephone surveys*, [w:] de Leeuw E.D., Hox J. J., Dillman D.A. (eds), International handbook of survey methodology, New York: Taylor & Francis Group.
- Steve K.W., Burks A.T., Lavrakas P.J., Brown K.D., Hoover J.B. [2008], *Monitoring telephone interviewer performance*, [w:] Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japac L., Lavrakas P.J., Link M.W., Sangster R.L. (eds), Advances in telephone survey methodology, Hoboken: John Wiley & Sons.
- Sykes W., Collins M. [1988], *Effects of mode of interview: experiments in the UK*, [w:] Groves R.M., Biemer P.P., Lyberg L.E., Massey J.T., Nicholls II W.L., Waksberg J. (eds), *Telephone survey methodology*, New York: John Wiley & Sons.
- Sztabiński F. [1995], *Kontrola w badaniach surveyowych: pytania i odpowiedzi*, „ASK. Społeczeństwo, Badania, Metody”, vol. 1/1995.
- Sztabiński P.B. [1999], *Wywiad telefoniczny ze wspomaganie komputerowym: nowa technika badawcza, czy wywiad kwestionariuszowy przez telefon?*, „Przegląd Socjologiczny”, 48(1).
- Tarnai J., Paxson M.C. [2005], *Interviewer judgments about the quality of activities that interviewers consider stressful, as well as to identify the strategies the interviewers use when they are in contact with difficult respondents*.
- Tucker C. [1983], *Interviewer effects in telephone surveys*, “Public Opinion Quarterly”, vol. 47(1).

Wojciech Jabłoński

#### DIFFICULT SITUATIONS IN TELEPHONE INTERVIEWS

##### Abstract

The results show that there are two types of duties causing stress to the interviewers: activities that may lead to an interview break-off and activities that cause cognitive difficulties to the interviewers. Methods adopted by the interviewers in dealing with problem respondents seem to be dependent on the type of troublesome situation. In instances when respondents experience difficulty in understanding the questions, most interviewers destandardize the interview protocol and simplify the vocabulary and syntax. Interviewers' reactions and strategies in other situations (i.e. reluctant, irritated and talkative respondents) are more complicated and are thus determined by many factors.

**Key words:** CATI, interviewers, telephone surveys, standardization.

SYLWIA MĘCFAL\*  
Uniwersytet Łódzki

## **PROBLEMY BADAŃ TERENOWYCH – WYBRANE KWESTIE METODOLOGICZNE, PRAKTYCZNE ORAZ ETYCZNE W BADANIU ZJAWISK „TRUDNYCH”**

### **Streszczenie**

Najbardziej interesujące zjawiska społeczne są na ogół trudno dostępne dla badacza. Trudności w tego typu projektach badawczych pojawiają się na każdym kroku: problem wiarygodności informatorów/badanych, problem wyboru odpowiedniej metody, problem niedostatecznej wiedzy na dany temat. Wszystko to powoduje, że badacz może znaleźć się w sytuacji, kiedy kontrola procesu badawczego jest niewielka. Czasami przeprowadzenie niektórych czynności badawczych może okazać się niemożliwe (np. ze względu na brak zgody na wejście do instytucji totalnej czy niechęć badanych do ujawniania informacji, które mogą uderzać w nich samych). Zjawiska trudno dostępne najczęściej opierają się metodom ilościowym. Złożoność takich zjawisk, ich specyfika, sprawiają, że badacze sięgają po metody jakościowe. Nie oznacza to jednak, że wybór ten chroni ich przed innymi niełatwymi decyzjami podczas pracy w terenie – może się okazać, że ostatnią instancją odwoławczą pozostaje intuicja badawcza i doświadczenie w pracy w terenie. Jak radzić sobie w terenie badając grupy zamknięte, wykluczone, specyficzne grupy zawodowe, społeczności lokalne, czy też zjawiska, które budzą wątpliwości etyczne: działania zakulisowe, korupcję, konflikt interesów? Jakich problemów możemy się spodziewać? W artykule postaram się zaprezentować przykłady zmagania badawczego ze zjawiskami trudno dostępnymi.

**Słowa kluczowe:** badania terenowe, obserwacja uczestnicząca, monografia terenowa, studium przypadku, jawność/niejawność procesu badawczego, zjawiska zakulisowe, problemy badań terenowych, triangulacja

---

\* sylwia.mecfal@wp.pl



## UZASADNIENIE TEMATU I WYBÓR TEKSTÓW

Badania terenowe dla początkującego badacza stanowią poważne wyzwanie. A. Trąbka: [2007: 81–101] bardzo trafnie, według mnie, porównuje wyprawę na badania terenowe do rytuału przejścia. Wiedzę zdobytą podczas badań terenowych można, jej zdaniem, porównać do wiedzy, jaką uzyskują podczas inicjacji nowicjusze. Taką wiedzę można zdobyć tylko drogą osobistego doświadczenia, jest to dla wielu badaczy moment przełomowy i niezwykle wartościowy [ibidem: 96].

Jeśli badania dotyczą zjawisk „trudnych” proces badawczy jest nawet bardziej skomplikowany. Tematyka badania zjawisk „trudnych” jest dla mnie istotna z kilku względów. Jako dość młody socjolog-badacz, choć doświadczony już wieloma różnorodnymi projektami badawczymi, często zastanawiam się nad wartością, jakością i poprawnością badań nad problemami, które mnie interesują najbardziej. Badania te dotyczą zjawisk społecznych znajdujących się na pograniczu rzeczywistości jawnej – dostępnej dla bezpośredniej obserwacji i rzeczywistości ukrytej – do której badacz nie ma bezpośredniego dostępu<sup>1</sup>. Decyzje badawcze, metodologiczne, przed którymi sama stanęłam, skłoniły mnie do tego, aby przyjąć się, jak z tematami „trudnymi” radzą sobie inni badacze. Swoje poszukiwania tekstów, które poruszałyby kwestie metodologiczne, etyczne czy też w końcu dawałyby praktyczne wskazówki badawcze, zaczęłam od zapoznania się z zestawieniem artykułów i książek o tematyce metodologicznej, które sporządziła Krystyna Lutyńska [2000a, 2002]. Zestawienia te zawierają publikacje poruszające tematykę metodologiczną, które ukazały się do 2000 roku. Brakowało w nich zatem tytułów najnowszych. Aby je odnaleźć przejrzałam najbardziej popularne czasopisma socjologiczne: „Studia Socjologiczne”, „Kulturę i Społeczeństwo”, „Przegląd Socjologii Jakościowej”, „Przegląd Socjologiczny”, a także periodyki „ASK” oraz „Analizy i próby technik badawczych w socjologii”. Dodatkowo czerpałam inspirację z publikacji książkowych, które poruszały tematykę zjawisk „trudnych” – takich, które były mi znane oraz takich, o których dowiedziałam się z Bibliograficznej Informacji Bieżącej wydawanej przez Polskie Towarzystwo Socjologiczne.

Lista tych materiałów na pewno nie jest wyczerpująca, jednak najważniejszym argumentem decydującym o tym, że brałam pod uwagę jakiś artykuł, jakąś

<sup>1</sup> Aktualnie na potrzeby swojej pracy doktorskiej prowadzę badania terenowe – studium przypadku, które zawiera w sobie różnorodne techniki zarówno wysoko standaryzowane jak i nisko standaryzowane. Tematyka pracy dotyczy problemów funkcjonowania prasy lokalnej w jej otoczeniu społecznym. Zjawiska przenikalności środowisk politycznego, biznesowego oraz dziennikarskiego, konflikt interesów to tylko kilka z tzw. zjawisk trudnych, które zamierzam opisać.

książkę czy monografię był fakt szerszej refleksji autora/autorów nad kwestiami metodologicznymi, etycznymi czy praktyką badawczą.

Mówiąc o „trudności” badań mam na myśli głównie trzy obszary:

1. Badania z zastosowaniem „trudnych” metod (inaczej metod złożonych czy skomplikowanych, do których zaliczam głównie metody jakościowe wiążące się z dłuższym czasem przebywania badacza w terenie, koniecznością bycia blisko badanych, czyli monografia terenowa, studium przypadku (*case study*), metoda etnograficzna, obserwacja uczestnicząca czy quasi-uczestnicząca).

2. Trudno dostępne zjawiska, problemy badawcze (należące do sfery zjawisk ukrytych ze względu na ich nielegalność, intymność, poczucie zagrożenia ze strony badanych, ale także ze względu na niezrozumienie i obcość jakiegoś świata dla badacza: np. przestępczość, korupcja, konflikt interesów, ale także seksualność, religia czy wykluczenie społeczne).

3. „Trudne” tereny badawcze (zatem miejsca, gdzie bardzo trudno prowadzi się badania, a niejednokrotnie jest to niemożliwe: instytucje zamknięte: więzienia, klasztory, seminaria duchowne, a także grupy wyznaniowe, sekty, grupy przestępcze, subkultury, grupy ekskluzywne).

Wyróżnienie tych trzech obszarów ma jedynie na celu pewne uporządkowanie kwestii, na których będę się koncentrować. W praktyce badawczej bardzo często te trzy obszary będą ze sobą współwystępować, bo przecież to złożoność problematyki i terenu badawczego decyduje o tym, że badacz wybiera takie metody, które najlepiej uchwycą specyfikę danego problemu (często są to właśnie te metody, które zaliczam tutaj do tzw. „trudnych”).

Ważnym zagadnieniem, które postaram się także zasygnalizować, jest według mnie to, jak prowadzić tak specyficzne badania, aby były one rzetelne, trafne i wiarygodne. Czyli: Jak zachować jakość w badaniach jakościowych? – jak głosi tytuł książki Uwe Flicka [2011]. Autor ten dodatkowo przytacza rozważania innego badacza, R. Bohnsacka [2005], który sformułował pewne tezy dotyczące standardów badań jakościowych. Chciałabym przytoczyć tu Tezę nr 1, która głosi „Metody i standardy badań jakościowych opracowuje się na podstawie empirycznej rekonstrukcji praktyk badawczych” [R. Bohnsack, 2005; za: U. Flick, 2011:41]. Zatem niezmiernie ważne dla każdego badacza powinno być opisywanie swoich praktyk oraz zapoznawanie się z praktykami stosowanymi przez innych badaczy. Pośrednie i niestety wycinkowe przedstawienie tych praktyk jest celem tego artykułu.

### JAK PROWADZI SIĘ BADANIA... NIEJAWNOŚĆ BADAŃ WŚ BADANIA W PEŁNYM ŚWIETLE

Badacze podejmując się trudnych tematów niejednokrotnie decydują się ukrycie swojej prawdziwej roli, bądź jedynie częściowo zdradzają cel swojej obecności w danym środowisku czy też podają fałszywy cel. Jack Douglas [1976; za Patton 1997: 163–200] twierdzi, że na ogół metody antropologicznych badań terenowych opierają się na założeniu, że ludzie są gotowi do współpracy, że pragną, by ich punkt widzenia był rozumiany i podzielany przez innych. On sam jednak uważa, że to „konflikt jest rzeczywistością naszego życia, podejrzliwość zaś jego wiodącą zasadą. [...] Jest to wojna wszystkich przeciwko wszystkim i nikt nie daje nikomu niczego za darmo, a przede wszystkim – prawdy [1976: 55; za: Patton, 1997: 171].

Ponadto, jak pisze Paweł Moczydłowski [1988: 49]: „Badacz powinien zdawać sobie sprawę, iż sama zauważalna obecność obcego w nieznannej mu społeczności może już ją zmienić. Powinien także pamiętać, że pomimo oswojenia jej ze sobą, może zmienić się ponownie, gdy pojawi się kolejny obcy.” P. Moczydłowski [1988, 1990: 23–37] przeprowadzając badania w więzieniu zdecydował się na ukrycie prawdziwego celu swoich badań. Uznał, że jest to jedyny sposób na to, aby jego badania przyniosły wiarygodne informacje. Dodatkowo przyjął on postawę, którą sam określał jako „naiwną” poznawczo [1990: 23–37], postawił siebie w roli ignoranta, który nic nie wie o życiu więziennym i nie przesądza o tym, czego się dowie. Każdy uczestnik życia więziennego był dla Moczydłowskiego kompetentnym i równoważącym informatorem. Ukrywanie prawdziwego celu badań czy też fałszowanie tego celu utrudnia badanym kontrolę nad rejestrowanymi przez badacza danymi [Moczydłowski 1990: 32], jednak wiąże się z różnymi trudnościami. Moczydłowski nie mógł prowadzić notatek w obecności badanych, nie mógł w żaden sposób rejestrować na bieżąco swoich obserwacji. Dodatkowo brak narzędzia badawczego spowodował, że bardzo szeroki był wachlarz zjawisk, które badacz musiał zapamiętać. Takie postępowanie może spotkać się z zarzutem subiektywizmu czy też z zarzutem braku możliwości intersubiektywnej kontroli, jednak jest to w wielu wypadkach jedyna droga uzyskania informacji o życiu w instytucji totalnej<sup>2</sup>.

<sup>2</sup> Za Goffmanem [1975, 2011] przyjmuję, że instytucja totalna to taka, gdzie całe życie mieszkańców toczy się w jednym i tym samym miejscu, podlega tej samej władzy. Przełamane są bariery, które oddzielają 3 dziedziny życia: miejsce spania, zabawy i pracy. Instytucja totalna ingeruje we wszystkie sfery życia mieszkańców. Cały ich dzień jest ściśle zaplanowany, także jedna czynność w wyraźnie przewidzianym czasie przechodzi w drugą. Plan narzucony jest z góry przez system

Decyzję o niejawności całego procesu badawczego podjął także Kamil Miszewski [2005: 65–93, 2007: 33–63], którego interesowała również specyfika życia więziennego. Sytuacja Miszewskiego była dodatkowo skomplikowana, ponieważ dokonywał on obserwacji uczestniczącej z pozycji więźnia<sup>3</sup>, a zatem nie obserwował tylko współwięźniów, ale także siebie samego. Ponadto cała sytuacja została wymuszona przez okoliczności obiektywne, a nie była początkowo przemyślanym i zaplanowanym działaniem badawczym. Ukrycie swoich badań było, według Miszewskiego, „warunkiem koniecznym dla zebrania [...] nieprzekształconych – przez sam proces badawczy – danych.” [2005:76]. Drugim powodem podjęcia decyzji o niejawności prowadzonego badania (oraz konieczność utrzymania tej niejawności) była obawa o osobiste bezpieczeństwo badacza. Dekonspiracja badacza-więźnia groziłaby sankcjami zarówno ze strony współwięźniów (pobicie, izolacja itp.) jak i ze strony administracji więzienia (odmowne załatwianie próśb i wniosków, przeniesienie, kierowanie do cel z ‘trudnymi’ więźniami itp.) [Miszewski 2005].

Podobnie jak w przypadku Moczydłowskiego, tak i w sytuacji K. Miszewskiego robienie notatek było bardzo utrudnione. W drugim przypadku jednak badacz nie miał możliwości wyjścia z badanej społeczności i zrobienia notatek po jednym dniu prowadzenia obserwacji. Musiał znaleźć sposób na to, aby prowadzić zapiski oraz jednocześnie nie wzbudzać większego zainteresowania. Początkowo robił notatki dość otwarcie, ale udawał, że są to notatki związane z jego studiami socjologicznymi (trzymał je w książkach socjologicznych wraz z innymi notatkami), później udawał, że pisze listy<sup>4</sup>. Widać zatem, że w przypadku sytuacji nietypowych wypracowane standardy na niewiele się zdają. Badacz musi podejmować decyzje sam, niejednokrotnie musi bardzo szybko reagować, a każde takie działanie może wpłynąć na przebieg procesu badawczego i na uzyskiwane rezultaty.

Technikę obserwacji ukrytej zastosowali także moi studenci prowadząc badania nad innym rodzajem instytucji, która ma niektóre cechy instytucji totalnej – seminarium duchownym<sup>5</sup>. Próby zbadania życia w seminarium duchownym za

---

formalnych rozporządzeń (regulamin), jego przestrzegania pilnuje zespół nadzorców. Występuje podział na 2 grupy: podwładni – jednostki żyjące wewnątrz instytucji, mające ograniczony kontakt ze światem zewnętrznym oraz personel – mniej liczna grupa, jednostki społecznie zintegrowane ze światem zewnętrznym. Każda z tych grup ocenia się na postawie wrogich stereotypów. Przemieszczenia między tymi grupami są bardzo ograniczone, istnieje duży dystans społeczny.

<sup>3</sup> Badacz przebywał dwa i pół roku w więzieniu, gdzie odbywał karę pozbawienia wolności orzeczoną za spowodowanie wypadku samochodowego.

<sup>4</sup> Szczegółowy opis dylematów badacza znajduje się w jego artykułach.

<sup>5</sup> Badania prowadzone były w ramach zajęć z metod badań społecznych latem i jesienią 2008 roku. Raport z badań jest dostępny na <http://eksoc.uni.lodz.pl/is/doc/seminarium.pdf>

pomocą technik jawnych (FGI z klerykami) nie powiodły się (nie uzyskaliśmy na nie zgody rektora seminarium duchownego), zatem jako koordynator badań podjęłam decyzję o zastosowaniu ukrytej obserwacji zewnętrznej podczas „dni powołaniowych” organizowanych w seminarium, podczas których każdy mógł przyjść do seminarium i zobaczyć, jak mieszkają klerycy oraz bezpośrednio z nimi porozmawiać. Obserwacja była co prawda krótkotrwała (zajęła kilka godzin), jednak udało się zebrać studentom bardzo ciekawy materiał (także o charakterze wizualnym – zdjęcia), który znacznie wzbogacił informacje uzyskane za pomocą innych technik (studenci przeprowadzali także wywiady kwestionariuszowe z księżmi oraz kilka wywiadów swobodnych z klerykami – jednak dostęp do tych drugich był dość trudny).

Innym przykładem konieczności ukrywania procesu badawczego jest badanie przeprowadzone przez Irenę Kleszcz [2004: 189–202], która wykorzystwała technikę obserwacji uczestniczącej do zdobycia informacji o stylu życia szarej strefy. Zastosowanie techniki jawnej w tego typu badaniu nie miałoby sensu. Informacje, do których chciała dotrzeć autorka należą do tzw. danych brudnych i ukrytych [Marx 2003: 7–50], zatem są to dane tajne i kompromitujące, których odkrycie byłoby dyskredytujące lub kosztowne dla kogoś (w tym przypadku dla badanych) w kategoriach różnego rodzaju sankcji negatywnych [ibidem: 8–9].

Badaczka, aby dotrzeć do interesujących ją kwestii rozpoczęła pracę w firmie, której jawna działalność była jedynie fasadą. Wejście do tej firmy nie było dla niej trudne, jak sama przyznaje, „odbyło się w sposób naturalny. Dzięki znanej mi osobie, która miała kontakty z interesującą mnie grupą i była dla tej grupy godna zaufania, udało mi się zatrudnić w (Spółce) [...]” [Kleszcz 2004: 195]. Jednak lojalność badaczki była sprawdzana na każdym kroku, stąd nie miała ona możliwości robienia notatek na bieżąco. Sporządzała je dopiero po powrocie do domu i, jak przyznaje, rekonstrukcja wszystkich zdarzeń była bardzo trudna i nie zawsze kompletna [ibidem: 199]. Inny sposób rejestracji czy też wykorzystanie innych technik badawczych nie było w tej sytuacji badawczej możliwe.

Zupełnie inne podejście do sposobu prowadzenia badań mają badacze, którzy sądzą, że cel badania nigdy nie powinien być tajemnicą, a samo badanie powinno być prowadzone w „pełnym świetle”. Anna Wyka [1990: 161–173, 1993] nawołuje do tego, żeby badacze przyjmowali postawę „poznawczo naiwną” i uznali podmiotowość badanych. To badany jest dla socjologów kompetentnym informatorem i zasługuje na to, żeby mieć pełne informacje o przedsięwzięciu badawczym. Badanie jest zatem traktowane przez autorkę jako wymiana. „Kontakt badawczy próbuje się budować na wzór opartej na wzajemności (zwykłej relacji międzyludzkiej). Porzuca się (hermeneutykę podejrzliwości), którą zastępuje

właśnie otwarta wymiana oparta na umowie, zaufaniu i szacunku wzajemnym. Badanie jako całość staje się (wspólną drogą), w której badani (znają swoje życie, my mamy swoje dyspozycje), badani (znają teren), a badacze (mają przewodnik), przy czym żadne z narzędzi nie jest (sztywne) i może być modyfikowane w trakcie procesu badawczego.” [Wyka 1990: 165–166].

Zatem to stanowisko badawcze porzuca paradygmat konfliktowy i koncepcję „śledztwa badawczego” [Douglas 1976; za: Patton 1997], a przyjmuje koncepcję Shilsa, który obserwację ukrytą czy prowadzenie niejawnego procesu badawczego uważa za „moralnie naganną manipulację” [1956; za: Patton 1997].

Taka idea przyświecała badaniom prowadzonym przez zespół profesora Andrzeja Sicińskiego w badaniach nad stylami życia. [Gliński 1990] Badanie było prowadzone metodą studium przypadku, w obrębie którego badacze stosowali różnego rodzaju jakościowe techniki zbierania materiałów. Założeniem całego procesu badawczego było zbliżenie się proponowanego przez Galtunga podejścia „dialogowego”, które zakłada jak najpełniejsze włączenie w proces badawczy całego zespołu badawczego oraz również samych badanych. Zespół Sicińskiego podjął próbę realizacji „idei swoistych badań aktywizujących (action research) polegających na przedstawieniu wstępnych wniosków i wyników badań osobom badanym, uzyskaniu opinii o tych wynikach, wspólnej dyskusji na ich temat prowadzonej pod kątem ich trafności oraz przydatności z punktu widzenia możliwych modyfikacji działań, zmian zachowań, dążeń życiowych itp.” [ibidem: 40] W przypadku niektórych badanych takie podejście nie sprawdziło się i ciągle traktowali badacze-socjologów jako „obcych”. Niemniej jednak zastosowanie takiej różnorodności technik i podejść w procesie badawczym przyniosło wiele wartościowych i interesujących informacji, które pozwoliły na stworzenie typologii stylów życia, opis tych stylów oraz odtworzenie zasad organizujących życie codzienne rodzin [ibidem:50].

W badaniach stylów życia badacze dążyli do tego, aby ich stosunki z badanymi uległy „prywatyzacji” [Łukasiewicz 1985: 99–115], aby byli postrzegani nie tylko jako badacze, ale również jako osoby prywatne. Sytuacja badawcza stawała się wówczas dla badanych podobna do zwyczajnych sytuacji życiowych, z jakimi spotykali się na co dzień, np. wizyta znajomego, sąsiada. Założeniem badaczy było jak najpełniejsze włączenie badanym w proces badawczy, zatem nie mogli ukryć tutaj swojej oficjalnej roli, stąd podjęli takie kroki, które pomogły tę oficjalność „oswoić”. Jak przyznaje Piotr Łukasiewicz [ibidem: 105] fakt, że badani godzili się na dalsze wizyty badaczy, był nie tyle efektem pełnego uświadomienia sobie przez badanych celu prowadzonych badań, a raczej było to przyzwolenie wynikające z sympatii do badacza jako osoby prywatnej. Z upływem czasu badani przywykli do wizyt badaczy, a nawet oczekiwali na te wizyty. Uzyskanie takich

prywatnych relacji z badanymi pozwoliło na zdobywanie informacji nie tylko podczas wywiadów swobodnych, ale również podczas swobodnych rozmów, gdzie badacze często uzyskiwali bardziej wartościowe i interesujące ich informacje. Prowadzenie badań „w pełnym świetle” pozwoliło zespołowi badawczemu na zdobycie zaufania badanych. Uznali oni, że „jedynie dłuższy, oparty na zaufaniu kontakt może przynieść jakościowo odmienne od dotychczasowych, sięgające głębiej dane o życiu badanych.” [ibidem: 112]

Grażyna Woroniecka [1999: 177–193] starała się pokazać, że stosowanie ukrytej obserwacji uczestniczącej w celu rozwiązania jakiegoś problemu społecznego, może znieść problem niejawności i nieetyczności takich działań. Badanie, które zaproponowała Woroniecka odbywa się na drugim planie, a „celem nawiązania i kontynuacji relacji jest rozwiązanie problemu.” [ibidem: 190] G. Woroniecka zastosowała modyfikację obserwacji uczestniczącej w celu zbadania struktury komunikacyjnej samorządu na przykładzie działania jednej ze wspólnot mieszkaniowych w Olsztynie. Posłużyła się realną sprawą, która stała do rozwiązania przed wspólnotą (zlecenie administrowania budynkiem za wynagrodzenie), aby ustalić reguły praktyki samorządowej. Sama, jako członek zarządu wspólnoty, była stroną w sprawie, co pozwoliło jej na ukrycie pod tą rolę roli obserwatora. Dzięki istnieniu tego problemu – sondy uczestnik-badacz konsekwentnie realizował swoją rolę formalną [w tym wypadku członka zarządu wspólnoty), przez co ukrycie roli badacza nie wzbudza już takich kontrowersji natury etycznej. „Jeżeli wskutek tego (sabotażu) socjologicznego dochodzi do konfliktu, to inicjatorem jest strona badana, która wybiera taką strategię komunikacyjną jako odchylenie od idealnego typu wzajemnej relacji.” [ibidem: 190] W swoim podejściu badaczka połączyła w pewnym sensie prowadzenie badania niejawnego ze strategią zaangażowania badanych w proces badawczy. Wprowadzenie problemu-sondy sprawiło, że cały proces badawczy znajdował się na drugim planie, a celem głównym stało się rozwiązanie problemu przez badanych oraz samego badacza-uczestnika tej grupy. Problem grupy był w tym wypadku także problemem badacza.

### **JAK BADAĆ ZJAWISKA NIEJAWNE, ZAKULISOWE?**

Zakulisowe wymiary życia społecznego<sup>6</sup> nie są popularnym wśród badaczy tematem. R. Sojak i D. Wincenty [2005: 13–67, 157–225] przeanalizowali dorobek polskich socjologów z lat 1989–2000 i udało się im odnaleźć jedynie

<sup>6</sup> Za Maciejem Gurtowskim i Janem Waszewskim [2009: 166–181] przyjmuję tutaj, że „zjawiska zakulisowe jako przedmiot badań obejmują takie działania społeczne, które są przez podejmujące je podmioty ukrywane w obawie przed sankcją.” Zamiar ukrycia takiego działania

sześć tekstów, które dotyczyły zakulisowych wymiarów życia społecznego oraz cztery dodatkowe publikacje, odnoszące się w ograniczonym stopniu do zjawisk niejawnych. Autorzy nie odnotowali w tym okresie ani jednej pracy doktorskiej czy habilitacyjnej, która dotyczyłaby tego typu zjawisk. Tymczasem to właśnie dotarcie do zjawisk niejawnych pozwala nam na pełniejsze zrozumienie rzeczywistości społecznej. Rolą socjologa jest zrozumienie własnego społeczeństwa oraz pomoc temu społeczeństwu, aby zrozumiało samo siebie [Ziółkowski 2000: 185; za: Sojak, Wincenty 2005: 18]. „Socjologia musi być zatem z natury „sztuką nieufności” i „sztuką podejrzeń”, powinna umieć powiedzieć coś ciekawego, interesującego i odkrywczego o społeczeństwie własnym [...], powinna być niepokorna, starać się przeniknąć fasadowe pozory, odsłonić rzeczywiste mechanizmy i zejść do poziomu czynników istotnych leżących zwykle pod powierzchnią zjawisk.” [ibidem: 18–19]. Bardziej zdecydowaną postawę wobec roli socjologii prezentuje tzw. socjologia demaskatorska (*muckraking sociology*), która bada nieprzystawanie oficjalnych deklaracji do realnie podejmowanych działań społecznych, zwłaszcza jeśli dotyczą one interesu publicznego. [Gurtowski, Waszewski 2009]

Dlaczego zatem tak mało jest prac dotyczących tej grupy zjawisk? Przyczyny takiego stanu rzeczy są zróżnicowane. Fakty dotyczące tego typu zjawisk należą na pewno do tych, które trudno ustalić (zatem są to zjawiska „trudne”). Ponadto status metodologiczny tych faktów można uznać za wątpliwy [Zybertowicz 2009: 159–160; za: Gurtowski, Waszewski 2009: 166]. Badacz może mieć trudności w zrozumieniu i właściwym zinterpretowaniu takich faktów, gdyż nie zawsze ma możliwość poznania kontekstu tych działań oraz zazwyczaj nie ma możliwości weryfikacji tych faktów. Informatorzy badacza są w takich sytuacjach przeważnie anonimowi, co może podważyć wiarygodność tak uzyskanych informacji. Ponadto badając tego typu zjawiska badacz nie może zwykle sięgnąć do doświadczeń innych badaczy, do pewnych standardów badania zjawisk zakulisowych, gdyż one po prostu nie istnieją. Czynnikiem dodatkowo zniechęcającym jest niepewność statusu takiego badania naukowego względem poniesionych kosztów [ibidem: 166]. Proces badawczy jest zwykle długotrwały, a badacz nigdy nie może być pewien tego, na ile wartościowe będą rezultaty tego przedsięwzięcia.

Ważnymi czynnikami mogą być w końcu te o charakterze psychologicznym i społecznym, jak zauważają Gurtowski oraz Waszewski [2009: 168]. Przecież badacz naruszając interesy innych osób i grup, naraża siebie na działania negatywne – mogą zdarzyć się pozwy sądowe, naruszenie dóbr osobistych czy dys-

---

nie musi być intencjonalny, choć stoją za nim pobudki racjonalne: uzyskanie przewagi organizacyjnej, rozszerzenie spektrum możliwych działań o takie, które są w danej społeczności zabronione.



kredytowanie badacza. Przeszkodą może być też bariera psychologiczna, opór przed „wejściem” w świat nieznany, nieprzewidywalny, wiążący się z dużymi kosztami emocjonalnymi.

I. Kleszcz [2004: 196–197], która prowadziła obserwację niejawną osób działających w szarej strefie, obawiała się o swoje bezpieczeństwo. Badani, sprawdzając jej lojalność, nigdy nie pozostawiali jej samej w biurze, wystawiali ją na próbę poprzez podkładanie cennych rzeczy do jej biurka i sprawdzanie reakcji, czy wreszcie opowiadali o tym, co stało się z poprzedniczką badaczki, kiedy ta nadużyła ich zaufania (zatem stosowali różnego rodzaju techniki zastraszania). Jak przyznaje autorka, w trakcie badania korzystała z porad psychologa w zakresie przebywania w środowisku stresogennym. Po zakończeniu badania autorka, obawiając się o własne bezpieczeństwo, pisywała pod pseudonimem.

Jak zatem badać zjawiska zakulisowe, aby przynajmniej ograniczyć te trudności, które wiążą się z takim przedmiotem badań? Maciej Gurtowski i Jan Waszewski [2009: 170–178] podają kilka rozwiązań metodologicznych, z których może korzystać badacz<sup>7</sup>. Zatem pomocne rozwiązania to: a) metafora „układanki”, a zatem rozpoczynanie badania od elementów, które można zidentyfikować w sposób łatwiejszy, znalezienie punktów odniesienia, które pozwolą nam na szybsze odtworzenie obrazu całego zjawiska; b) metoda białego wywiadu, czyli systematyczne gromadzenie informacji pochodzących ze źródeł ogólnie dostępnych, ich analizę oraz przedstawianie wyników (przykładowe źródła to: Internet, prasa i inne mass media, dokumentacja np. firm i przedsiębiorstw, dostępne rejestry, decyzje sądowe, wydawnictwa marketingowe); c) umiejętne docieranie do informacji publicznie dostępnych, zdolność przewycięzania biurokratycznej inercji; d) informatorzy; e) umiejętne identyfikowanie dezinformacji; f) metoda obserwacji uczestniczącej.

Do techniki obserwacji uczestniczącej sięgnęła wspomniana wcześniej badaczka Irena Kleszcz. Także Marcin Spławski [2004: 132–141] oparł swoje dociekania o tę technikę. Celem autora była próba opisu funkcjonowania władzy

<sup>7</sup> Opisane są one w artykule autorów znacznie szerzej. W tym miejscu omówię je jedynie skrótowo. Dodatkowo autorzy opisują kilka heurystyk, z których warto skorzystać: a. dziennikarskie 5xW+H (a zatem podstawowe dla dziennikarza pytania: *who, what, where, when, how and why*), b. siedem złotych pytań kryminalistyki (co, gdzie, kiedy, jak, jaką metodą (jakim narzędziem) oraz kto); c. metodologia NIK (4xN – ustalenie działań nielegalnych, niegospodarnych, niecelowych oraz nierzetelnych); d. korzystnie z bogatej literatury dotyczącej działalności tajnych służb (*intelligence studies*); e. teoria wyboru publicznego (racjonalność działań i nakierowanie na zysk może sugerować, że potencjalnych sprawców należy szukać wśród tych, którzy odnieśli korzyść); f. teoria zмовы kartelowej; g. wskaźniki inferencyjne.

samorządowej w kontekście działania lokalnych grup interesu. Wybór tej techniki podyktowany był dwoma głównymi powodami: brak dobrej orientacji w środowisku lokalnym bardzo utrudnia identyfikację grup interesów; środowisko polityczne jest dość opornie poddaje się próbom eksploracji, zwłaszcza w sytuacji, gdy badacz interesuje nie tylko strona formalna i jawna, ale także sfera nieformalna i zjawiska niejawne czy zakulisowe. Przyjmując rolę radnego, badacz miał możliwość zaobserwowania procesu decyzyjnego od wewnątrz, a także uzyskał dostęp do głównych aktorów na interesującej go scenie samorządu lokalnego (a zatem możliwość przeprowadzenia nieformalnych rozmów z politykami, biznesmenami czy urzędnikami). Uzyskane dane z obserwacji, nieformalnych rozmów, a także z analizy materiałów prasowych pozwalają zaliczyć postępowanie badawcze jako *case study*. Te zróżnicowane materiały pozwoliły M. Spławskiemu na identyfikację najważniejszych grup interesu w miejscowości oraz na wykazanie, że „lokalne grupy interesu, względnie trwałe i zinstytucjonalizowane, odznaczają się daleko posuniętą tendencją do działania poza oficjalnymi kanałami artykulacji. Korzystniejsze dla nich wydają się dojścia nieformalne związane z używaniem brudnej formy kapitału społecznego<sup>8</sup>”. [Spławski 2004: 140]

Studium przypadku jako główną metodę w swoim procesie badawczym zastosował także Bent Flyvbjerg [2005: 41–71]. Według niego studium przypadku jest szczególnie przydatne w identyfikowaniu „czarnych łabędzi”. Ma tę właściwość „ze względu na pogłębione podejście do problemu: to, co wydaje się (białe), po bliższym przyjrzeniu się często okazuje się (czarne)”. [ibidem: 51] Autor swojego „czarnego łabędzia” znalazł, gdy przeprowadzał swoje pierwsze studium przypadku dotyczącego polityki miejskiej i planowania miasta Aalborg w Danii. Okazało się, że w działaniach pomiędzy sferą biznesu, polityki i administracji nie ma miejsca na konkurencję i wolny rynek oraz na demokrację przedstawicielską. Jak przyznaje autor: „lokalni przedsiębiorcy stanowili silną siłę, która negocjowała nielegalne układy z politykami i administracją, służące blokowaniu konkurencji i wolnego rynku i dające im uprzywilejowaną pozycję”. [ibidem: 51]

Widać zatem, jak „trudne” dla badacza metody pozwalają mu uchwycić takie aspekty zjawisk, do których nie mógłby dotrzeć stosując metody standaryzowane i pewne gotowe wzory postępowania badawczego. To, co dla badacza może być trudnością, a zatem brak wypracowanych metod badania zjawisk zakulisowych

---

<sup>8</sup> Brudny kapitał społeczny, to według rozmienia M. Spławskiego [2004: 135–136], sytuacja, „w której jednostki lub grupy o wysokim poziomie kapitału społecznego wykorzystują go do niegodnych celów.” „Brudny kapitał społeczny w przeciwieństwie do jego klasycznej wersji nie jest produktywny: nie pomnaża zasobów szerszym zbiorowościom lub społeczeństwom, lecz je drenuje.”

czy zjawisk niejawnych, może stać się zaletą podczas pracy w terenie. Badacz nie jest ograniczony, jest bardziej elastyczny i może łatwiej przystawać się do sytuacji, która się pojawia. Zachowany zostaje tzw. kontekst odkrycia – jeśli posłużyć się terminologią stosowaną przez metodologię teorii ugruntowanej.

Zatem wbrew temu, co sugeruje cytowany wcześniej A. Zybertowicz, niewypracowane wzory badań zjawisk zakulisowych czy niejawnych, nie wydają się stanowić bariery w badaniu takich zjawisk. Zdaniem B. Flyvbjerga badacze przypadków zakładają „zglębianie zjawisk na własną rękę w miejsce stosowania ich map. To, co E. Goffman [1963] nazywa (kulisami) społecznych zjawisk, musi również podlegać badaniu, tak jak boczne uliczki, o których mówił Wittgenstein.” [2005: 63]

Każdy badacz, który decyduje się na pomijanie zjawisk niejawnych czy też zadowala się docieraniem do sfery jawnej, powinien mieć świadomość, że stanowią one nieodłączną część rzeczywistości społecznej. Pomijanie ich powoduje trywializację przedmiotu badań socjologii, a także docieranie do zafałszowanego, niepełnego obrazu rzeczywistości, co stanowi wyraz negacji wielowymiarowości świata społecznego [Gurtowski, Waszewski 2009: 169].

### **GDY SOCJOLOG TO „OBCY”**

Każdy socjolog podejmujący się badań terenowych początkowo (a czasami podczas całego badania) traktowany jest przez badanych i sam też czuje się jak „obcy”. Świadomość takiej sytuacji wpłynęła na postępowanie zespołu profesora Sicińskiego i jego zespołu, kiedy podejmowali oni próbę „prywatyzacji” swojej relacji z badanym [Łukasiewicz, op. cit.], aby badani przestali traktować ich tylko jako badaczy, ale też dostrzegli w nich osoby prywatne i nieco przyzwyczaili się do nietypowej dla nich sytuacji badania. Dla badacza fakt akceptacji i obdarzenia go zaufaniem także był istotną kwestią. Jednak sytuacja badań sfery, która jest w pewnym sensie bliska badaczowi (style życia rodzin miejskich) i może on do pewnego stopnia utożsamić się z sytuacją badanych czy też postawić się w ich roli, nie przystaje w żaden sposób do sytuacji badawczej, kiedy badacz wchodzi w bardzo specyficzny świat badanych i próbuje zrozumieć funkcjonujących w nim badanych. Badacz jest tam przecież tylko „wizytatorem” [Rakowski 2002: 67–79], nie doświadcza takiego życia na co dzień.

Tomasz Rakowski [2002] opowiada o zjawisku „nędzy” jak o kulturowo obcym sposobie życia ludzi. Zdaniem autora, w badaniach nad ubóstwem nie jest możliwa do spełnienia podstawowa zasada badań antropologicznych, czyli „autorytet doświadczenia”. Rakowski określa badacza jako „Cudzoziemca”,

który ma jedynie możliwość „wizytacji nędzy”, gdyż każdy pobyt w terenie to tylko pobyt chwilowy, a badacz ostatecznie zawsze wraca, wraz z materiałami, które zdobył, do miejsca, z którego pochodzi. [ibidem: 67] „Świat ubóstwa” jest zatem dla badacza niemal jak badanie innej kultury, a fakt, że badacz nie jest elementem tej kultury sprawia, że jego możliwości poznawcze i interpretacyjne są ograniczone. Dla badacza doświadczenie takiego „obcego” świata wiąże się często z nieprzydatnością przyjętych założeń, gdyż najczęściej „nie jest tak, jak przyjmowaliśmy”. „Każde spotkanie jest przyjęciem wyzwania, próbą stanięcia wobec całkiem obcej rzeczywistości, niezrozumiałego człowieka, wobec kultury, której decydującym aspektem ‘jest jej nędza’.” [ibidem: 78]

O poczuciu „inności” i badaniu ludzi „żyjących inaczej” wspomina Elżbieta Tarkowska [1996] zastanawiając się nad problemami badań, które prowadziła z zespołem: „Życie codzienne w domach pomocy społecznej”. Badanie miało na celu opis warunków życia i sposobów zaspokajania potrzeb osób niepełnosprawnych żyjących w domach pomocy społecznej. „Inność” świata badanych wyrażała się choćby w przebywaniu w środowisku zamkniętym, w instytucji, którą E. Tarkowska zaliczyła do grupy instytucji totalnych. Badani zatem to osoby żyjące cały czas w jednym miejscu, niemal zupełnie odizolowane od świata zewnętrznego i jego problemów. Dodatkowo, byli to szczególnego rodzaju „inni”, jak określa ich E. Tarkowska [1996: 19–28]. Były to osoby niepełnosprawne, często chore, co powodowało, że były one zupełnie zależne od innych: od personelu pracującego w domu pomocy społecznej, a także, w trakcie prowadzonych badań, od badacza. Jak zauważa badaczka, dodatkowym problemem dla członków zespołu badawczego był fakt, że nie posiadali oni dostatecznej wiedzy medycznej, żeby zrozumieć do końca sytuację badanych. Musieli polegać na personelu medycznym, który niejako częściowo tłumaczył im świat i sytuację badanych, a raczej własne rozumienie tej sytuacji.

Anna Hańcza [1996: 29–45] na podstawie swoich doświadczeń podczas prowadzenia długoletniej obserwacji na wsi wskazuje, że również obecnie można zauważyć odmienności kulturowe między osobami pochodzącymi z dużych miast (zwłaszcza gdy posiadają one wyższe wykształcenie) a mieszkańcami wsi. Te odmienności dotyczą stylu i sposobu życia, obyczajów tradycji, języka; traktowania określonych wartości oraz zachowań związanych z tymi wartościami (traktowanie choroby, dzieci, stosunek do kobiet, niepełnosprawnych); stosunku do „obcych” czy „świadomości” psycho-społeczno-kulturowej („godność osobista i grupowa”). [ibidem: 34] Jeśli badacz trafi w takie „odmienne” środowisko, może to powodować u niego poczucie „inności”. Jak przyznaje autorka, od momentu osiedlenia się w miejscowości, którą obserwowała, starała się, aby jej stosunki

z mieszkańcami były dobre. Jednak nie udało jej się uniknąć zachowań niezręcznych. Mieszkańcy wsi traktowali ją tak, jak innych przybyszy z miasta, których dzielili na „działkowiczów”, „turystów”, „letników” oraz „gości”. [ibidem: 35] Aby uzyskać miano „dobrego sąsiada”, które według badaczki jest bliskie znaczeniowo pojęciu „swojskości”, starała się udzielać w zakresie wzajemnej pomocy, czyli poprzez wyświadczanie mieszkańcom tej miejscowości różnych przysług. A. Hańcza nazwana została „prawdziwie dobrą sąsiadką” przez jednego z gospodarzy, gdy pożyczyła mu mały telewizor, aby mógł obejrzeć mecz, kiedy jego własny się popsuł [ibidem: 41]. Fakt prowadzenia obserwacji uczestniczącej, a zatem dłuższego przebywania w społeczności, zdaniem autorki, wpłynął na zmniejszenie się jej poczucia „inności” oraz umożliwił zdobycie informacji szerszych, bardziej pogłębionych, takich, których nie zdobyłaby zapewne z pozycji „innego”.

Problem może sprawić identyfikowanie badacza nie tylko jako „innego” czy „obcego”, ale też błędne identyfikowanie badacza jako pracownika instytucji, lekarza, urzędnika, a zatem ogólnie rzecz biorąc osobę, która może coś zmienić w sytuacji badanego.

Podczas badań na temat warunków życia mieszkańców Łodzi i funkcjonowania instytucji pomocy społecznej, gdzie badanymi byli klienci pomocy społecznej<sup>9</sup> kilka razy zetknęłam się z taką błędną identyfikacją. Badani oczekiwali, że udzielone w wywiadzie informacje zmienią coś w sposobie funkcjonowania instytucji pomocy społecznej i w rozdziale środków finansowych, który ich zdaniem nie był sprawiedliwy. Prośby o „zrobienie czegoś” były kierowane jednak bezpośrednio do mnie, mimo moich tłumaczeń, że nie jestem przedstawicielem instytucji pomocy społecznej. Czasami błędna identyfikacja mojej osoby jako przedstawiciela instytucji pomocy społecznej powodowała jednak zupełnie inną reakcję – tzw. blokadę informacji. Jedna z rodzin, dopiero po kilkunastu minutach wywiadu zaczęła szczerze opowiadać o swojej sytuacji, kiedy po raz kolejny dałam im zapewnienie, że jest to badanie naukowe, a nie kontrola tego, czy faktycznie należą im się środki przyznane z pomocy społecznej. Podobne sytuacje, które miały miejsce podczas badania biedy w rodzinie, opisuje Krystyna Lutyńska [2000b: 199–233]. Badani także mieli tendencję do identyfikacji badacza jako pracowników „opieki społecznej” czy też innej instytucji nadrzędnej nad gminnym ośrodkiem pomocy społecznej. Ta błędna identyfikacja powodowała (podobnie jak w opisywanym przeze mnie wcześniej przypadku) ukrywanie przez

<sup>9</sup> W badaniach uczestniczyłam w ramach ćwiczeń terenowych jeszcze w czasie studiów w 2003 roku.

badanych swojej sytuacji materialnej. Badani nie zmieniali swoich identyfikacji, mimo prawidłowego przedstawienia się badaczy. [ibidem: 208].

Sytuację błędnej identyfikacji badacza przedstawia także Katarzyna M. Borucka [1992: 145], gdzie ankieter przeprowadzający wywiad z osobami chorymi na raka często był utożsamiany z pracownikiem medycznym, czy też nawet samym lekarzem. Zdaniem autorki, wywiady były prowadzone w bardzo specyficznym i trudnym dla badanych momencie – jakim niewątpliwie jest leczenie poważnej choroby, czym można tłumaczyć te błędne identyfikacje.

W swoim doświadczeniu badawczym dostrzegłam także zachowania badanych, które można odebrać jako sprawdzanie badacza oraz jako próbę sprowadzenia sytuacji nietypowej do „codziennej” wizyty sąsiada czy kogoś z rodziny. Podczas wywiadów z klientami pomocy społecznej badani, chcąc wyrazić swoją gościnność i pozytywne nastawienie do sytuacji, oferowali od zwyczajnej szklanki wody do własnych „kapci”. Odrzucenie takiej oferty mogłoby wydać się niegrzeczne i zakłócić dalszy przebieg wywiadu. Podobne odczucia, w jeszcze bardziej złożonej sytuacji badawczej, opisuje Izabela Ślęzak [2009: 181–199] zajmująca się zjawiskiem prostytutki. Badane prostytutki podczas wywiadów były szczególnie wrażliwe na objawy braku akceptacji, pogardy czy odrazy ze strony badaczki. Dużą wagę przykładają do tego, czy autorka zaakceptuje propozycję herbaty, poda im rękę czy wreszcie usiądzie na łóżku, które zwykle służy do pracy. [ibidem: 192]

Powyższe rozważania pokazują, że zarówno identyfikacja badacza jako „obcego”, jego autoidentyfikacja jako „innego”, jak i błędna identyfikacja roli badacza mogą w znacznym stopniu wpłynąć na rezultaty prowadzonych badań i jakość materiałów, które on zbierze. Badacz powinien dbać o relacje z badanymi i o to, aby prawidłowo odczytywali oni jego intencje. Niemniej ważnym zadaniem jest jednak samoświadomość badacza i umiejętność zachowania się w zaskakujących sytuacjach, których podczas badań terenowych może zdarzyć się wiele. To prawda, że badacz czasami jedynie „wizytuje” jakiś świat społeczny i nie jest w stanie do końca go zrozumieć, ważne by był tego świadomy i traktował te wizyty jako możliwość własnego rozwoju.

### **RÓŻNORODNOŚĆ I WIELOŚĆ MATERIAŁÓW A RZETELNOŚĆ I TRAFNOŚĆ WNIOSKÓW**

Metody, które traktuję w tym artykule jako „trudne” mają ze sobą wiele wspólnego. Dostarczają badaczowi wielu, bardzo różnorodnych materiałów,

niekiedy bardzo szczegółowych, które czasami mogą przekraczać możliwości analityczne jednego badacza. Dodatkowo wobec tych materiałów często wysuwane są zarzuty dotyczące ich wiarygodności, a rezultatom pochodzącym z analiz takich materiałów zarzuca się niski stopień rzetelności, trafności czy też brak możliwości budowania wniosków ogólnych.

Postępowaniem, dzięki któremu zdaniem wielu badaczy, można lepiej i głębiej zrozumieć dane zjawisko oraz podnieść wiarygodność danych uzyskiwanych w badaniach terenowych jest triangulacja (mowa o danych zbieranych głównie metodami jakościowymi, choć nie tylko – Alan Bryman [2006] pisze o powodach stosowania triangulacji w kontekście łączenia wyników z badań prowadzonych metodami jakościowymi i ilościowymi).

Można wyróżnić kilka rodzajów triangulacji: a. triangulację danych (korzystanie z kilku różnych źródeł danych w ramach jednej metody), b. triangulację badaczy (badania przeprowadzane są przez różnych obserwatorów czy prowadzących wywiady, aby ograniczyć wpływ indywidualnych preferencji badaczy), c. triangulację teoretyczną (wykorzystanie wielu ujęć teoretycznych do zinterpretowania uzyskanego materiału). d. triangulację metodologiczną (zatem zastosowanie różnorodnych metod zbierania danych w celu przezwyciężenia słabości poszczególnych metod) [Denzin 1970, za: Flick 2011: 82–87].

Denzin w swoim początkowym ujęciu kwestii triangulacji traktował ją jako postępowanie umożliwiające walidację wyników uzyskiwanych w badaniach terenowych, podnoszenie ich rzetelności oraz stwarzanie podstaw dla rozwoju teorii [ibidem: 86]. Dyskusja, jaka przetoczyła się nad jego koncepcją sprawiła jednak, że zmienił on nieco swoje stanowisko i w późniejszych pracach widział triangulację jako postępowanie dające możliwość „głębszego zrozumienia danego zagadnienia”, niekoniecznie zaś powodującego uzyskanie bardziej trafnych czy obiektywnych danych [Flick 2011: 91].

Argumenty stojące za stosowaniem triangulacji są jednak bardzo różnorodne. A. Bryman [2006: 104–107] przytacza wyniki analizy zawartości, której celem było odtworzenie uzasadnień stosowania triangulacji metodologicznej podawanych przez badaczy.<sup>10</sup> Udało mu się zaklasyfikować te uzasadnienia według następujących kategorii:

- a) uzyskanie większej trafności;
- b) przezwycięzenie słabości metod i wykorzystanie ich mocnych stron;

---

<sup>10</sup> Tekst A. Brymana odnosi się do łączenia metod ilościowych i jakościowych, jednak sądzę, że większość z argumentów podnoszonych przez autora można odnieść także do łączenia metod jakościowych.

- c) uzyskanie bardziej kompletnego oraz wyczerpującego opisu zjawiska;
- d) uzyskanie danych o charakterze procesualnym;
- e) możliwość uzyskania odpowiedzi na różnorodne pytania badawcze;
- f) wyjaśnianie rezultatów uzyskanych jedną metodą przez drugą;
- g) uzyskanie zaskakujących wyników, które można wyjaśnić używając drugiej metody;
- h) przygotowanie narzędzi badawczych;
- i) dobór jednostek badania lub przypadków do badania – uzasadniony rezultatami osiągniętymi z użyciem innych metod;
- j) uzyskanie bardziej wiarygodnych wyników;
- k) uzyskanie lepszego opisu kontekstu zjawiska;
- l) materiał ilustracyjny;
- ł) większa użyteczność wyników;
- m) tworzenie i weryfikowanie hipotez w ramach jednego projektu badawczego;
- n) różnorodność perspektyw (uzyskanie także punktu widzenia badanych);
- o) uzyskanie „mocniejszych” wyników, wzmacnianie wyników uzyskanych jedną metodą przez wyniki uzyskane przez stosowanie innej metody.

Strategia triangulacji jest nieodłączną częścią metod, wokół których koncentrują się moje rozważania w tym artykule. Jeszcze zanim samo pojęcie triangulacji zostało przeniesione na grunt nauk społecznych, jako cechy charakterystyczne techniki obserwacji uczestniczącej, monografii terenowej, studium przypadku czy metod etnograficznych były wskazywane stosowanie zróżnicowanych źródeł danych, a także różnorodnych technik i sposobów zbierania materiałów w celu stworzenia jak najpełniejszego opisu zjawiska (o czym pisali choćby Jan Szczepański w 1950 r.; za: Sosnowski 1985: 125–139; czy później Jan Lutyński [1961] w swoich rozważaniach nad antropologiczną monografią terenową).

Warto przyjrzeć się także charakterystyce metody etnograficznej, jaką prezentuje Michael Angrosino [2010: 45–46]. Zdaniem autora, wyróżnia się ona od innych sposobów prowadzenia badań naukowych, gdyż opiera się na pracy w terenie, ma charakter spersonalizowany (badacz ma codzienny bezpośredni kontakt z badanymi, jest zarówno uczestnikiem jak i obserwatorem zdarzeń), jest wieloczynnikowa (zawiera w sobie różnorodne techniki badawcze – „tak, by za pomocą triangulacji sformułować wniosek, którego pewność rośnie, właśnie dzięki wykorzystaniu różnych dróg dotarcia do niego”), wymaga długotrwałego zaangażowania badacza, ma charakter indukcyjny, opiera się na dialogu (badani mają możliwość komentowania na bieżąco wniosków i interpretacji wyciąga-



nych przez badaczy), ma charakter holistyczny (badacz dąży do uzyskania jak najpełniejszego obrazu badanego zjawiska czy badanej grupy).

Na gruncie socjologii polskiej częściej na określenie tego rodzaju badań używa się pojęcia badań monograficznych i często odnosi się je do badań społeczności lokalnych. Monografia terenowa to właśnie taka procedura badań, która opiera się na bezpośrednim kontakcie badacza z badanymi, w której badacz zbiera informację za pomocą zróżnicowanych technik badawczych<sup>11</sup>. [Kwaśniewicz 1993: 153–160, 1999: 43–50; Lutyński 1961: 36–59, 1994] Dodatkowo zastosowanie tzw. monografii problemowej, gdzie za punkt wyjścia bierze się problem istotny dla danej społeczności [ibidem: 154; ibidem: 41], pozwala badaczowi na stworzenie bardziej spójnego obrazu zjawiska (ogranicza się także tzw. eklektyczność wniosków, co często jest zarzutem pod adresem tego typu opracowań).

Takie ujęcie badań terenowych powoduje jednak, że badacz staje przed ogromnym wyzwaniem, wysiłkiem intelektualnym oraz organizacyjnym. Ponadto, musi liczyć się z tym, że jego wysiłek nie przez wszystkich badaczy może być doceniony. Musi liczyć się z tym, że pod adresem wniosków sformułowanych na podstawie takich badań będą kierowane różnego rodzaju zarzuty: brak możliwości uogólniania, brak możliwości poddania procesu badawczego intersubiektywnej kontroli (ewentualnie bardzo utrudniona), przez co podważana może być rzetelność zebranych materiałów oraz trafność wyprowadzanych wniosków.

Z takimi głosami spotkał się Bent Flyvbjerg [2005], kiedy zdecydował się na przeprowadzenie pogłębionego studium przypadku, które miało pomóc autorowi w zrozumieniu tego, jak władza i racjonalność oddziałują na siebie i jak kształtują one środowisko, w którym żyjemy. Jak przyznaje autor wielokrotnie powtarzane i przeczytane uwagi krytyczne sprawiły, że on sam nabrał dystansu wobec metody studium przypadku, jednak pomysłu na badanie nie porzucił. Przeprowadził analizę tekstów naukowych dotyczących tej metody, która pozwoliła mu stwierdzić, że zarzuty formułowane pod adresem metody mają charakter nieporozumień (czy też uproszczeń). W swoim artykule badacz najpierw formułuje pięć nieporozumień dotyczących metody studium przypadku, a następnie kolejno je koryguje. Te nieporozumienia to:

Nieporozumienie nr 1: Ogólna teoretyczna (niezależna kontekstowo) wiedza jest bardziej wartościowa od wiedzy konkretnej, praktycznej (uwarunkowanej kontekstowo).

---

<sup>11</sup> Nie tylko technik jakościowych, ale także ilościowych oraz analiz danych zastanych: dokumentów osobistych czy dokumentów urzędowych.

Nieporozumienie nr 2: Ponieważ nie można dokonywać uogólnień na podstawie studium przypadku, to studium przypadku nic nie wnosi do rozwoju nauki.

Nieporozumienie nr 3: Studium przypadku jest najbardziej użyteczne dla stawiania hipotez, to jest w pierwszym etapie procesu badawczego, podczas gdy inne metody są bardziej użyteczne w testowaniu hipotez i budowaniu teorii.

Nieporozumienie nr 4: Studium przypadku zawiera skrzywienie skierowane ku weryfikacji, gdyż sprzyja skłonności badacza do potwierdzania przyjętych wcześniej założeń.

Nieporozumienie nr 5: Na podstawie specyficznych studiów przypadku trudne jest często podsumowanie i rozwinięcie ogólnych twierdzeń i teorii. [ibidem: 43–44]

Już na samym wstępie autor stwierdza, opierając się na tezie sformułowanej przez T. Kuhna, że „dyscyplina naukowa bez dużej liczby porządnie opracowanych studiów przypadku nie dostarcza systematycznie przykładów empirycznych, a dyscyplina bez przykładów jest nieefektywna.” [ibidem: 41]. Odnosząc się do kolejnych „nieporozumień” przedstawia argumenty pokazujące, że studium przypadków to metoda, która pozwala nam zdobywać wiedzę o świecie społecznym, której nie uzyskalibyśmy innymi metodami, pozwala nam znajdować „czarne łabędzie” tam, gdzie inni widzą tylko te białe, oferuje możliwość zdobycia wiedzy konkretnej, zamiast daremnego poszukiwania uogólnień (co na gruncie nauk społecznych zdarza się rzadko). Ponadto autor pokazuje, że na podstawie studium przypadku częściej zdarza się falsyfikować nasze hipotezy dotyczące badanego zjawiska niż je potwierdzać. „Jeżeli za punkt wyjścia przyjmujemy proces nabywania wiedzy, to zrozumiemy, dlaczego badacz przeprowadzający studium przypadku często kończy uwalniając się od założonych pojęć i teorii. [...] Prostsze formy rozumienia muszą ustępować bardziej złożonym, gdy ktoś staje się z początkującego badacza ekspertem.” [ibidem: 60].

Nie ma tutaj miejsca, aby szczegółowo odnosić się do wszystkich argumentów B. Flyvbjerga, jednak już na podstawie tych kilku stwierdzeń widać, że część zarzutów formułowana pod adresem tej metody jest niezasadna i wynika raczej z braku wiedzy i doświadczenia w tym zakresie<sup>12</sup>.

<sup>12</sup> Autor odnosząc się do swoich własnych doświadczeń z pobytu w Uniwersytecie Harvarda wspomina argumenty przytaczane przez tamtejszych badaczy: „zapomnij o potocznej mądrości, weź się do pracy i zacznij robić studia przypadków” [ibidem: 45].

### PROBLEMY NATURY ETYCZNO-MORALNEJ – ZAMIAST ZAKOŃCZENIA

Dylematy etyczne są nieodłączną częścią prowadzenia badań terenowych. To, że kwestie i problemy etyczne w wielu przypadkach dyskutowane są w artykułach badaczy prowadzących badania terenowe, świadczy o ich wysokim stopniu samoświadomości i dużym namyśle nad specyfiką przedmiotu badań socjologa.

Gdy badacz podejmuje się eksploracji zjawisk trudnych, czy przeprowadzania badań w instytucjach, które na co dzień bronią dostępu do wielu informacji, musi zdawać sobie sprawę, że jego postępowanie będzie budzić wątpliwości, bardzo często będzie wywoływać w nim samym poczucie łamania pewnych reguł czy wykorzystywania nieświadomości badanych.

W omawianych przeze mnie przykładach badań niejawnych badacze stykali się z dwoma rodzajami dylematów. Z jednej strony, nie czuli się jak badacze czy obserwatorzy, a raczej jak „podglądacze”. Zwłaszcza w relacji I. Kleszcz [op. cit.] tego typu problem jest wyrażany przez autorkę wprost, a jej dylematy stają się jeszcze większe wraz z „wsiąkaniem” w grupę, nawiązywaniem znajomości z badanymi. Druga strona tych dylematów, to koszty, jakie ponosi sam badacz, podejmując się tego typu badań. W relacjach [Kleszcz; Miszewski – op. cit.] pojawia się obawa badaczy o swoje bezpieczeństwo (przy badaniach szarej strefy czy badaniach w więzieniu), ale też rysują się koszty społeczne takich przedsięwzięć (w końcu badacz na dłuższy czas zostawia swój własny świat, żeby lepiej poznać świat badanego). I. Kleszcz [op. cit.] wyraża przekonanie, że w sytuacji prowadzenia badań niejawnych, badacz jest w ciągłym napięciu (obawa o odkrycie prawdziwej roli), może to prowadzić do rozchwiania emocjonalnego i niebezpieczeństwa autodestrukcji badacza (autorka zmuszona była do korzystania z pomocy psychologa).

Problem emocjonalnego angażowania się badacza dotyczy jednak także badań jawnych, prowadzonych w pełnym świetle. Członkowie zespołu badawczego zajmujący się stylami życia [Gliński; Łukasiewicz – op. cit.] dążyli wręcz do tego, aby wytworzyć sobie z badanymi relacje prywatne, chcieli zakamufłować w ten sposób swoją formalną rolę – badacza. Jednak wywoływało to w nich niepokój, iż zgoda badanych na obecność badaczy w domach wynikała nie z tego, że badani byli świadomi celu i przydatności prowadzonych badań, a z tego, że zwyczajnie prywatnie polubili badacza. Badacze mieli zatem poczucie swoistego manipulowania badanymi [Łukasiewicz op. cit.], a przez to również zatracania idei prowadzenia badań, jaką na wstępie przyjęli.

O swoich dylematach w związku z nawiązywaniem prywatnych, przyjacielskich relacji z badanymi pisze także Jolanta Lisek-Michalska [1996: 199–205]. W badaniach życia codziennego w domach pomocy społecznej (badania o charakterze monograficznym, *case study*) autorka przeżywała konflikt pomiędzy rolą badacza a wewnętrznym nakazem do działania i niesienia pomocy. Jak przyznaje, wielokrotnie zastanawiała się nad tym, jaka jest jej rola: badacza stojącego z boku zdarzeń, czy też uczestnika sytuacji. Pojawiały się także dylematy dotyczące tego, co jest ważniejsze: uzyskanie informacji czy też podjęcie działania w sprawie i zaburzenie sytuacji badawczej.

Badacze często też podnoszą problem badania cierpienia [Borucka op. cit.] czy też „cudzego nieszczęścia” (jak określa to E. Tarkowska op. cit.). Często wiąże się to z badaniem ludzi zależnych (niepełnosprawnych fizycznie czy psychicznie), którzy czasami nie są w stanie udzielić badaczowi świadomej zgody na badanie. O problemie tym pisze Jakub Niedbalski [2009: 199–217], który przeprowadzał obserwację uczestniczącą w domach pomocy społecznej przeznaczonych dla osób upośledzonych umysłowo: „Pomimo otwartej i jawnej postawy obserwatora [...], badacz pozostawał po prostu „nierozpoznawalny”, co było związane z ograniczonymi możliwościami percepcyjnymi oraz trudnościami interpretacyjnymi osób niepełnosprawnych intelektualnie.”

Wreszcie ostatnią rzeczą, o której chciałabym w tym miejscu wspomnieć, to kwestia „znużenia pracą w terenie” [Gliński op. cit.: 42]. Badania terenowe to proces długotrwały, wymagający od badacza dużego nakładu sił psychicznych i fizycznych. W trakcie badań badacz napotyka różnorodne osoby i sytuacje, z którymi pewnie nie spotkałby się, gdyby nie wszedł w świat badanych. Nie zawsze te spotkania będą dla badacza pozytywnym impulsem do prowadzenia dalszych działań, zdarzą się też sytuacje czy osoby nieprzyjemne. P. Gliński [ibidem: 42] podpowiada, że w takich sytuacjach przydatne okazały się być materiały mające porządkować prowadzone badanie: a zatem instrukcje, dyspozycje do rozmów z badanymi, tematy do obserwacji. Nie ograniczały one sposobu prowadzenia badań czy pola obserwacji, a raczej przypominały sedno prowadzonego badania a często pomagały w rozwiązaniu sytuacji trudnych.

Na poruszone wyżej problemy nie ma lepszych i gorszych odpowiedzi. Badacz musi dostosować swoje postępowanie do sytuacji, z jaką ma do czynienia. Istnieją oczywiście standardy prowadzenia badań, którymi czasami może posiłkować się badacz. Jednak nie zawsze są one przydatne w konkretnym problemie, z którym styka się badacz. Zjawiska „trudne”, „trudne” tereny badawcze często wymykają się takim standardom, dlatego warto zapoznawać się z doświadczeniami innych badaczy i choćby pośrednio szkolić swój warsztat. Nic jednak nie zastąpi praktyki

„terenowej”. Jak podkreśla I. Kleszcz [op. cit.: 201]: „długookresowe badanie z zastosowaniem ukrytej obserwacji uczestniczącej [...] może skutecznie nauczyć dyscypliny w pracy badawczej, praktycznego działania w terenie, pozwala na budowanie własnego warsztatu badawczego, co dla początkującego badacza jest bardzo istotne.”

## BIBLIOGRAFIA

- Angrosino Michael [2010], *Badania etnograficzne i obserwacyjne*, Wydawnictwo Naukowe PWN: Warszawa.
- Borucka Katarzyna M. [1992], *Specyfika wywiadów kwestionariuszowych z chorymi na raka*, Analizy i próby technik badawczych w socjologii, t. IX, IFiS PAN.
- Bryman Alan [2006], *Integrating quantitative and qualitative research: how is it done?*, Qualitative Research vol. 6, Sage Publications, London, Thousand Oaks and New Delhi.
- Flick Uwe [2011], *Jakość w badaniach jakościowych*, Wydawnictwo Naukowe PWN: Warszawa.
- Flyvbjerg Bent [2005], *Pięć mitów o badaniach typu studium przypadku*, Studia Socjologiczne nr 2.
- Gliński Piotr, *Założenia i praktyka jakościowych badań stylów życia*, Kultura i Społeczeństwo nr 1, 1990.
- Goffman Erving [1975], *Charakterystyka instytucji totalnych*, [w:] *Elementy teorii socjologicznych*, (red) Derczyński W., Jasińska-Kania A., Szacki J., PWN: Warszawa.
- Goffman Erving [2011], *Instytucje totalne. O pacjentach szpitali psychiatrycznych i mieszkańcach innych instytucji totalnych*, GWP: Sopot.
- Gurtowski Maciej, Waszowski Jan [2009], *Redukcja do jawności. O pomijaniu zjawisk zakulisowych w badaniach socjologicznych*, [w:] Bąk Aneta, Kubisz-Muła Łukasz (red.), *Metody, techniki i praktyka badań społecznych*, Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej w Bielsku-Białej: Bielsko-Biała.
- Hańcza Anna [1996], „Swoi” i „obcy” w badaniach na wsi, Kultura i Społeczeństwo nr 1.
- Kleszcz Irena [2004], *Wykorzystanie ukrytej obserwacji uczestniczącej w badaniu stylu życia szarej strefy*, Kultura i Społeczeństwo nr 2.
- Kwaśniewicz Władysław [1993], *Badania monograficzne*, Kultura i Społeczeństwo nr 3.
- Kwaśniewicz W. [1999], *O potrzebie rewitalizacji badań monograficznych. Nowe wyzwania dla socjologii akademickiej*, ASK. Społeczeństwo, Badania, Metody 8.
- Lisek-Michalska J. [1996], *Relacja badacz-badany – problemy metodologiczne i etyczne (na marginesie badań nad życiem codziennym w domach pomocy społecznej)* Acta Universitatis Lodzensis Folia Sociologica.
- Lutyńska Krystyna [2000a], *Metodologia badań socjologicznych: bibliografia adnotowana księzek za lata 1945–1999*, Wydaw. IFiS PAN: Warszawa.
- Lutyńska Krystyna [2000b], *Bieda w rodzinie. Psychologiczny, społeczny i kulturowy kontekst wywiadów z przedstawicielami trzech pokoleń*, [w:] Tarkowska E. [red.], *Zrozumieć biednego*, Typografia: Warszawa.

- Lutyńska Krystyna [2002], *Metodologia badań socjologicznych. Cz. 2, Bibliografia artykułów metodologicznych w wybranych czasopismach oraz książkach za lata 1945–2000*, Wydaw. IFiS PAN: Warszawa.
- Lutyński Jan [1961], *Antropologiczna monografia terenowa i badania społeczno-kulturowych przeobrażeń w Polsce współczesnej*, Przegląd Socjologiczny, vol. XV/2.
- Lutyński Jan [1994], *Antropologiczna monografia terenowa i badania społeczno-kulturowych przeobrażeń w Polsce współczesnej*, [w:] Lutyński J., *Metody badań społecznych. Wybrane zagadnienia*, ŁTN: Łódź.
- Łukasiewicz Paweł [1985], *Metoda dialogowa w praktyce badawczej. Na przykładzie badań stylu życia*, [w:] Lutyńska K. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, IFiS PAN: Warszawa.
- Marx Gary T. [2003] *Uwagi na temat odkrywania, gromadzenia i oceny ukrytych i brudnych danych*. ASK. Społeczeństwo, Badania, Metody 12.
- Miszewski Kamil [2005], *Socjolog w kryminale: więzienie jako nieprzyjazny teren badań*, Studia Socjologiczne, nr 3.
- Miszewski Kamil [2007], *Kiedy badacz jest tajnym agentem. O postrzeganiu niejawnego obserwacji uczestniczącej jako etycznie problematycznej, metodach badań ilościowych, zakulisowych wymiarach życia społecznego i ich związku ze wszystkim tym, o czym przed chwilą*, Przegląd Socjologii Jakościowej, T. III, nr 2, pobrany kwiecień 2011 [http://qualitativesociologyreview.org/PL/archive\\_pl.php](http://qualitativesociologyreview.org/PL/archive_pl.php).
- Moczydłowski P. [1988], *Drugie życie w instytucji totalnej*, Warszawa.
- Moczydłowski Paweł [1990], *O sposobach wglądu w sekrety stosunków międzyludzkich. Przypadek instytucji totalnych*, Kultura i Społeczeństwo nr 1.
- Niezbalski Jakub [2009], *Niepełnosprawność intelektualna w perspektywie badań społecznych. Rozważania nad zastosowaniem metod jakościowych*, [w:] Bąk Aneta, Kubisz-Muła Łukasz (red.), *Metody, techniki i praktyka badań społecznych*, Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej w Bielsku-Białej: Bielsko-Biała.
- Pattin Michael Q. [1997], *Obserwacja – metoda badań terenowych*, [w:] Leszek Korporowicz (red.), *Ewaluacja w edukacji*, Warszawa: Oficyna Naukowa.
- Rakowski Tomasz [2002], *Antropolog jako wizytator nędzy. Rozważania o kulturze niemocy*, Kultura i Społeczeństwo nr 4.
- Skrzypczak Adam [2009], *Łączenie źródeł i triangulacja jako strategie poprawy jakości badań empirycznych*, [w:] Bąk Aneta, Kubisz-Muła Łukasz (red.), *Metody, techniki i praktyka badań społecznych*, Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej w Bielsku-Białej: Bielsko-Biała.
- Sojak Radosław, Wicenty Daniel [2005], *Zagubiona rzeczywistość. O społecznym konstruowaniu niewiedzy*, Oficyna Naukowa: Warszawa.
- Sosnowski Adam [1985], *Obserwacja uczestnicząca. Osiągnięcia i perspektywy*, [w:] K. Lutyńska (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, IFiS PAN: Warszawa.
- Spławski Marcin [2004], *Partykularyzm lokalny. O polskim samorządzie terytorialnym z perspektywy lokalnych grup interesu*, [w:] Żuk Piotr (red.), *Demokracja spektaklu? Kondycja polskiego życia publicznego 15 po zmianie systemowej*, Scholar: Warszawa.
- Ślęzak Izabela [2009], *Metodologiczne problemy badań nad zjawiskami „trudnymi” na przykładzie prostytutki*, [w:] Bąk Aneta, Kubisz-Muła Łukasz (red.), *Metody, techniki i praktyka*

- badan społecznych*, Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej w Bielsku-Białej: Bielsko-Biała.
- Tarkowska Elżbieta [1996], *Socjolog w instytucji totalnej. Uwagi na marginesie badań*, Kultura i Społeczeństwo, nr 1.
- Trąbka Agnieszka [2007] *Wyprowa na badania terenowe jako rytuał przejścia. Analiza treści dzienników antropologicznych*, Kultura i Społeczeństwo nr 2.
- Woroniecka Grażyna [1999], *Elementy obserwacji uczestniczącej w badaniach nad samorządem terytorialnym*, [w:] Rostocki W.A., Lutyńska K., Domański H. (red.), *Spojrzenie na metodę. Studia z metodologii badań socjologicznych*, IFiS PAN: Warszawa.
- Wyka Anna [1990], *Ku nowym wzorom badań społecznych w Polsce. Cechy badań jakościowych w ostatnich latach*, Kultura i Społeczeństwo, nr 1.
- Wyka Anna [1993], *Badacz wobec doświadczenia*, Wyd. IFiS PAN: Warszawa.

Sylwia Męcfal

**PROBLEMS IN FIELD WORK/FIELD STUDIES – CHOSEN METHODOLOGICAL, PRACTICAL AND ETHICAL ISSUES IN EXPLORING “DIFFICULT” PHENOMENA**

Abstract

The most interesting social phenomena are usually difficult to access for a researcher. Different problems in such research projects appear every step a researcher take: reliability and credibility of informants, choosing a right research method, having insufficient knowledge about the subject. All these factors can cause that a researcher has not enough control over research process. Sometimes a researcher might not even be able to conduct everything he/she planned (e.g. he/she will be refused to conduct research in a total institution or the informants will not be willing to share pieces of information which are harmful for themselves). Quantitative methods are usually of no use in exploring such phenomena. Their complexity, specifics cause that researchers use qualitative methods. Choosing these methods does not mean that a researcher can avoid difficult decisions whilst doing the field work – it can happen that the last source of reference is researcher's intuition and experience in field work. What are the ways of conducting research on closed or excluded groups, total institutions, specific professional groups, local communities or the phenomena that cause ethical doubts: actions 'behind the scenes' or other covert activity involving corruption or conflict of interest? What are the difficulties and problems we can expect? In my paper I present examples of different research on phenomena that are difficult to investigate.

**Key words:** field work, participant observation, ethnographic research, case study, hidden/open research process, 'behind the scenes' phenomena, problems in field work, triangulation.

RAFAŁ MACIĄG  
Uniwersytet Warszawski

## SPOŁECZNA KONSTRUKCJA DOKUMENTÓW I DANYCH URZĘDOWYCH. DANE MEDYCZNE

### Streszczenie

Konstrukcja społeczna oznacza, że dokument lub dane są wytworem organizacji, w której powstają. Dokument jest na tyle wiarygodny, na ile organizacja pragnie wykazać swoje osiągnięcia albo ukryć wady. W ten sposób powstają sprawozdania urzędowe, kościelne lub bankowe. Jednym z rodzajów danych są dane medyczne: samobójstwa, dane psychiatryczne czy dane na temat aborcji. W artykule poruszono problem małego zaufania socjologów do danych urzędowych i jednocześnie większej akceptacji administracyjnych danych ilościowych. Artykuł zwraca uwagę na koncepcje fenomenologiczne i etnometodologiczne, przyczyniające się jednocześnie do rozwoju analizy jakościowej w socjologii. Szczególnym przykładem może być rejestracja samobójstw i badanie definicji sytuacji w ustalaniu danych pierwotnych.

Urzędowy charakter danych i dokumentów jako wytworu instytucji wiąże się z odrębnym traktowaniem procesu społecznego, jakim jest etap zbierania i przygotowania danych oraz ich publikacji. Jednocześnie koncepcja obiektywności danych, którą badacz przyjmuje, jest koncepcją rozumienia nauki, ponieważ – w ujęciu fenomenologiczno-konstrukcyjnym – dokumenty i dane są traktowane *per se*, jako samodzielne przedmioty badania. Taki proces badawczy warto analizować w momencie powstawania danych, to jest w chwili ich gromadzenia w organizacji lub zbierania materiałów do sprawozdania.

**Słowa kluczowe:** dane medyczne, dokumenty urzędowe, analiza jakościowa.

---

\* raf.maciag@gmail.com



### CO TO JEST KONSTRUKCJA SPOŁECZNA DANYCH URZĘDOWYCH?

W socjologii dane urzędowe wykorzystywane są zwykle jako materiał do prowadzenia badań ilościowych. Wśród zagadnień metodologicznych mają one znaczenie ze względu na porównywalność wskaźników społecznych, rejestrację zdarzeń i ocenę wiarygodności źródeł. Już w klasycznej socjologii dane urzędowe stanowiły podstawę weryfikacji twierdzeń teoretycznych w *Samobójstwie* Émile'a Durkheima [Durkheim 2006]. Socjologowie rozwijający koncepcje interakcyjne, fenomenologiczne i etnometodologiczne uznali jednak, że należy badać sytuacje społeczne, w których powstają dokumenty i dane wewnątrz organizacji. W tym kontekście „konstrukcja społeczna” oznacza statystykę lub dokument, będący wytworem urzędów i instytucji publicznych. Według Hammersleya i Atkinsona [Hammersley, Atkinson 2000] konstrukcja społeczna, np. dokument lub dane są faktami społecznymi, poddawany interpretacji badacza. A. Sułek [2002] pisze, że dokumenty urzędowe stanowią produkt procedur społecznych, dla socjologa mogą być samodzielnym przedmiotem badań [por. Sułek 2002: 128–129].

Termin „konstruowanie społeczne” wprowadzili do socjologii P. Berger i T. Luckmann [Berger, Luckmann, 1966/1983] na oznaczenie postrzegania świata w sposób subiektywny w interakcjach opartych na społecznym wytwarzaniu rzeczywistości. Jednostki wytwarzają i interpretują znaczenia społeczne w procesie życia codziennego. Opisy dokumentów i danych sporządzano od lat 60. ubiegłego wieku za pomocą obserwacji sytuacji społecznych i odpowiedniego zestawu pojęć [np. Blumer 2007; Glaser, Strauss 2009]. W tym ujęciu konstrukcjami społecznymi są dane, przy czym termin „dane” stał się założeniem, polegającym na korzystaniu przez badaczy z gotowych zestawień ilościowych. Aaron V. Cicourel [1964] zauważa, że w ten sposób socjologowie przyjmują ustaloną kulturowo rzeczywistość społeczną, na ogół nie sprawdzając i nie analizując wcześniejszego procesu społecznego, w którym powstają dane ankietowe podczas badań terenowych. Takie podejście odnosi się również do dokumentów i sprawozdań urzędowych. Uwagi swe odnosił do zachowań występujących podczas gromadzenia danych w czasie ich zapisywania. Cicourel [1968] opisał szczegółowe wnioski z badań w sądzie dla nieletnich, obserwując pracę funkcjonariuszy, którzy tworzyli akta i stosowali procedury formalne w celu umieszczenia „przestępstw” w różnych kategoriach. Proces negocjacji z nieletnimi obejmował podział zachowań na zgodne i niezgodne z normami społecznymi, ale dopuszczał też kategoryzowanie na pochodzenie z różnych klas społecznych. I tak jeżeli „przestępca” był recydywistą, okazywał się dewiantem, ale jeżeli

przydarzało mu się lekkie wykroczenie, wtedy darowano mu karę i zwykle był lepiej traktowany przez policję.

Ze względu na zainteresowania socjologiczne dane urzędowe podzielić można na kilka rodzajów. Jednym z nich są dane o chorobach psychicznych, inny przykład stanowią dane zbierane przez policję na temat przestępstw, jak wyżej, a także inne stanowiące problemy społeczne, na przykład: ubóstwo, bezrobocie, wypadki samochodowe, alkoholizm, zachorowania na AIDS. W socjologii dane te mogą stanowić podstawę analiz porównawczych, lecz jako dane społeczne z punktu widzenia analiz interakcji, są danymi wytwarzanymi w organizacjach, i mogą być analizowane przy wykorzystaniu metod jakościowych [Hammersley, Atkinson 2000].

Termin „dane urzędowe” (ang. *official statistics*) oznacza ogólnie dane wytwarzane przez urzędy i generalnie organizacje biurokratyczne, następnie publikowane w sprawozdaniach głównie ze względu na cele finansowe dla instytucji nadzorujących. W polskim języku prawno-administracyjnym (i języku badań statystycznych) obowiązuje termin „dane statystyczne” (por. *Ustawa o statystyce publicznej z 1995 r.*).

Wraz z dostrzeżeniem analiz interakcji wzrosło znaczenie badania problemów społecznych w socjologii [por. Frysztański 2009]. Perspektywa konstruktywizmu społecznego opiera się na analizie problemów społecznych, wytwarzanych w najbliższej przestrzeni społecznej jednostek, grup i instytucji społecznych. W poruszonym kontekście znaczenie ma interpretacja danych wynikających na przykład ze sporów wokół aborcji, w których głównymi argumentami są liczby, przedstawiane z punktu widzenia określonej instytucji. Tego typu działania są interpretowane w zależności od przekazu i intencji odbiorcy, a wielkość danego zjawiska jest podawana za pomocą „obiektywnych” danych, wiarygodnych dla nadawcy. Wiarygodność i jej ocena jest elementem świata racjonalnego i paradygmatu normatywnego w nauce, który jest odległy stanowisku zarówno fenomenologów jak i konstruktywistów.

Dokumenty i dane analizowane są coraz częściej jako element etnografii czy też antropologii organizacji [Hammersley, Atkinson 2000; Silverman 2008]. W socjologii jakościowej zestawienia ilościowe traktowane są jako część badania kultury organizacyjnej, a ich rozumienie to wynik analiz retoryki i analiz lingwistycznych dokumentów oraz badań języka używanego w instytucji. Takie badanie to pewien proces, w którym dane są zbierane, przygotowywane, przeliczane i publikowane przez instytucję. Dane są gromadzone w toku negocjacji znaczeń społecznych [Taylor 1982], a w takim procesie instytucja nierzadko uwikłana

jest w problemy społeczne (religijne czy polityczne), od których często zależy społeczne definiowanie i stąd „wiarygodność” i wartość danych.

Od czasu, kiedy przedmiotem analiz w socjologii stał się świat pisany, problemami badawczymi są język, nazewnictwo dokumentów, treść danych, a nawet poszczególne zmienne i klasyfikacje statystyczne, występujące w służbie zdrowia (np. ochrona życia) czy w polityce społecznej (np. bezrobocie, zarządzanie organizacją, analizy wynagrodzeń). Te kwestie są przedmiotem sporów społecznych i politycznych, ale także metodologicznych przy gromadzeniu danych urzędowych.

### SPÓLECZNE ZNACZENIE SAMOBÓJSTW

Dyskusja o samobójstwach w kontekście pomiaru zła moralnego rozpoczęła się w XIX wieku. Pomysł na teoretyczne potraktowanie tematu i metodę pomiaru miał É. Durkheim [2006], który w swoim studium wyjaśnił problem natężenia samobójstw. Za pomocą „obiektywnych” danych statystycznych sprawdzał teorię i w ten sposób wyjaśniał wewnętrzną naturę spoistości zbiorowości, a samobójstwa nazwał faktami społecznymi. Dopiero nurt fenomenologiczny zaczął dyskusję nad szczegółami wiarygodności takiego pomiaru i w ogóle nad przebiegiem całej rejestrowanej sytuacji społecznej. Jack D. Douglas [1967] określił społeczne znaczenie samobójstwa jako konstruowanie społeczne. Douglas jak i inni socjologowie z nurtu fenomenologicznego, a także etnometodologowie, nie podzielali koncepcji funkcjonalnych o integracji społecznej, zdając się na obserwację sytuacji codziennych i powstawania danych pierwotnych.

Na dowód słabej jakości danych urzędowych Douglas wskazuje najpierw na analizy Maurice’a Halbwachsa [1930], ucznia Durkheima, który pisał o problemie wiarygodności źródeł administracyjnych na przełomie XIX i XX wieku. Halbwachs podaje przykład reformy statystyki urzędowej na terenie Prus po zjednoczeniu państw niemieckich w roku 1871. Warto podać tę argumentację, ponieważ ukazuje ona naturę problemów gromadzenia danych. Otóż reforma polegała na przejęciu obowiązków administracyjnych od kościołów przez państwo. Zmiana wykazała różnice w rejestracji samobójstw: odnotowano wzrost z 4984 do 6171 [Halbwachs za: Douglas 1967: 193]. W ramach tej samej reformy w Saksonii odnotowano wzrost samobójstw z 745 w 1876 r. do 1114 w 1877 r. [Tamże]. Przyczyną tego stanu było rejestrowanie danych przez księży katolickich i duchownych protestanckich, zbierających dane o zgonach, wśród których nie uwzględniano Żydów. Podobne zmiany miały miejsce w Cesarstwie Austro-

-Węgierskim. Istotna była również rewizja pojęcia samobójstwa, mająca teraz świecki charakter, ponieważ według nowszej formuły zdarzenia odnotowywano w karcie zgonu i przekazywano je do kartoteki urzędowej a nie jedynie do ksiąg kościelnych.

W zależności od tego, jakie znaczenie nadadzą śmierci urzędnicy lub osoby bliskie zmarłego, taką też nazwę przyjmuje się w karcie zgonu. Dzieje się tak z powodów religijnych i obyczajowych kultury zachodniej, niedopuszczającej niektórych faktów wprost, a lekarz stwierdzający zgon zostaje do tej wersji przekonany. Według Douglasa jednostki konstruują wiedzę o samobójstwie w społecznie znaczącej sytuacji *post factum* zdarzenia, czyli w procesie negocjacji znaczeń, wydając różne sądy o przyczynie śmierci. Douglas zajmuje stanowisko, iż samobójstwo jako zjawisko społeczne staje się konstrukcją w wyniku społecznej definicji sytuacji, w której jest ono określane jako „wypadek” lub „nieszczęście”. Fakt zakończenia życia nie jest poddawany w wątpliwość, ale przyczyna śmierci może być różnie podawana w zależności od sytuacji społecznej, a w definiowaniu jej bierze udział kilka osób: rodzina zmarłego, lekarz, policjant, koroner<sup>1</sup>. Odnosząc się do sytuacji sobie współczesnej Douglas uważa, że wskaźniki samobójstw są błędnie zapisywane u źródeł, ponieważ śmierć samobójcza traktowana bywa przez lekarzy jako wypadek, wynik choroby lub śmierć naturalna. Fakt ten wynika zarówno z niewiedzy, lecz także z celowego działania innych osób, którym zależy na nierozpowszechnianiu prawdziwej wersji zdarzenia. Podejście interakcyjne nie bierze pod uwagę oceny *wiarygodności* źródła, ale raczej opis definicji sytuacji i jej subiektywny kontekst. „Fakty społeczne” zapisane w statystyce i tak bywają wątpliwe i nie da się ich do końca ustalić, a zastany kształt danych wpływa na dalszą dyskusję dotyczącą przyczyn śmierci. W prawie polskim i według urzędowych kryteriów klasyfikowane fakty muszą być zgodne z normami i regułami panującymi w prokuraturze i na policji oraz według wytycznych statystyki, gdzie rejestruje się samobójstwa i próby samobójcze [por. Jarosz 2004].

Bliska tej koncepcji wydaje się socjologia rozumiejąca Webera [2002], zakładająca znaczące działania i interpretacje ich sensów („motywów”) w konkretnych sytuacjach społecznych [Douglas 1967: 235–239]. Proponuje on analizy działań społecznych przy pomocy metody empatii i rekonstrukcji wydarzeń, subiektywnie odtwarzanych przez jednostki. W ten sposób wyjaśnianie zachowań samobójczych opiera się na definicjach tworzonych w sytuacjach społecznego ustalania przyczyn śmierci, w których bierze się pod uwagę okoliczności podawane przez rodzinę

---

<sup>1</sup> Koroner jest urzędnikiem angielskim odpowiedzialnym za prowadzenie śledztwa w sprawie nagłej śmierci; w Polsce – lekarz medycyny sądowej.

i warunki życia danej osoby uznanej za samobójcę. Autor proponuje badanie za pomocą metody biograficznej, aby ustalić motywacje do odebrania sobie życia i wskazuje na psychiatrów, mających doświadczenie w pracy z pacjentami po próbach samobójczych. Metodę biograficzną można krytykować za ograniczenia polegające na rekonstrukcji zdarzeń i za stroniczą interpretację podawanych zeznań, jednak charakteryzuje się ona pozyskiwaniem określonych, wprawdzie nielicznych, tropów w ustalaniu przebiegu zdarzenia: historii życia, wywiadów pogłębionych z bliskimi czy możliwości prowadzenia rozmów z rodziną. Tak więc podejście propagujące społeczne zdefiniowanie sytuacji wywarło także wpływ na ocenę wiarygodności wskaźników urzędowych samobójstw.

Steve Taylor [1982], uwzględniając wnioski poprzednika, przedstawił tradycję Durkheimowską i reakcję na nią ze strony socjologii humanistycznej i fenomenologicznej, i uogólnił stanowiska badaczy zajmujących się samobójstwami do czterech rodzajów. Z powodu traktowania przez badaczy danych urzędowych owe cztery stanowiska Taylor nazywa „typami idealnymi”, i omawia względy teoretyczne i metodologiczne. Literatura na ten temat rzeczywiście jest szeroka, a tutaj wypada jedynie nadmienić, iż pierwsze stanowisko zajmuje typ ograniczonej akceptacji wobec wskaźników samobójstw, ponieważ bywają one przydatne jako ogólna miara zjawiska, lecz są poddawane krytyce w chwili porównywania na poziomie międzykulturowym i/lub międzynarodowym z powodu szczegółowych różnic w rejestracji (istnieje oczywiście standaryzacja porównań, ale nie o to chodzi w szczegółowych rozróżnieniach). Drugie stanowisko zakłada pełną akceptację danych porównawczych i całkowite zaufanie do rzetelności źródeł (przykład analiz szkoły Durkheima). Trzecie stanowisko charakteryzuje się sceptycyzmem do danych, ponieważ badacze mają wątpliwości co do procedur rejestracyjnych i definicji urzędowych (zbliżone do stanowiska pierwszego). Stanowisko czwarte reprezentuje badaczy odrzucających wskaźniki samobójstw, przyjmujących rozwiązania proponowane przez fenomenologów i etnometodologów. Taylor daleki jest od jednoznacznej krytyki danych ilościowych, a na gruncie wyrażanych wątpliwości do wskaźników samobójstw stwierdza, że rejestracja przyczyn śmierci jako samobójstwa bywa wątpliwa z powodu upływu czasu od chwili zdarzenia oraz okoliczności, w której lekarz wypisujący kartę zgonu umieszcza informację taką, jaką otrzymuje o danej przyczynie. Wiedza ta wiąże się ściśle z kwestią wiarygodności źródeł i problemem analizy danych, ich porównywalnością i rzetelnością. Taylor zwraca uwagę, że w analizowanych przypadkach społeczno konstruowania samobójstw najbardziej sceptyczni wobec danych ilościowych są psychiatrzy, wskazujący na rozbieżności między kryteriami dowodów podawanymi przez koronerów a kwalifikacjami zgonów

umieszczanymi w szpitalnych zestawieniach przez lekarzy, wpisującymi je dla celów statystycznych i płatniczych [Taylor 1982].

Wypada jeszcze dodać, że B. Pescosolido i R. Mendelsohn [1986] przeprowadzili badania wskaźników ilościowych, których wynik zaprzeczył koncepcji społecznej konstrukcji samobójstw. Oczywiście badania te nie brały pod uwagę socjologii życia codziennego i nie przeprowadzono analiz etnograficznych lub innych. Sprawdzili oni, czy samobójstwa są prawidłowo rejestrowane, a jeśli tak, to w jaki sposób można wpływać na wyniki wnioskowania teoretycznego. Badanie wykazało, że statystyka umieralności była prowadzona nierzetelnie i zawierała m.in. wiele błędów losowych w klasyfikowaniu przyczyn samobójstwa. Dlatego postawiono wniosek o występowaniu społecznego konstruowania wskaźników samobójstwa, negatywny dla testów teorii socjologicznych. Jednak dla potwierdzenia wniosku przeprowadzono dodatkowe badanie na niezależnych zbiorach danych, tym razem polegające na sprawdzeniu występowania błędów losowych wskaźników samobójstw. Okazało się, iż dane te nie miały wpływu na zmienne społeczne w testowaniu socjologicznych teorii samobójstwa.

## RODZAJE DANYCH URZĘDOWYCH

Warto teraz spojrzeć na cel badań ilościowych i ich rodzaje oraz odpowiedzieć na pytanie, dlaczego socjologowie korzystają z nich tylko w wybranych dziedzinach. Z punktu widzenia organizacji biurokratycznej i metod badań dane dzielą się na dwa rodzaje: pierwotne i wtórne. Dane pierwotne są rejestrowane bezpośrednio ze źródeł (np. w komórkach organizacyjnych zakładów pracy, uczelniach, szpitalach i innych). Tam są wypełniane formularze statystyczne i przesyłane do ośrodków podliczających dane. Badacz na ogół nie ma dostępu do takich danych, ponieważ najpierw podlegają one agregacji, czyli procesowi grupowania ich w zestawienia i obliczaniu średnich i wskaźników. Dlatego badacz korzysta z drugiego rodzaju danych – z danych wtórnych, czyli z zestawień i zmiennych. Następnie dane udostępniane są jako wyniki zbiorcze instytucji. Dane urzędowe są to dane zastane w rocznikach, biuletynach statystycznych, w archiwach i dokumentach urzędowych. Dane wtórne można podzielić z kolei na wewnętrzne materiały przedsiębiorstw, szpitali lub uczelni oraz na dane zewnętrzne, czyli opracowania statystyczne lub historyczne, przeznaczone do celów edukacyjnych lub do celów zarządczych firm czy urzędów.

Celem prowadzenia urzędowych badań statystycznych jest porównywalność wskaźników społecznych i ekonomicznych oraz zarządzanie finansami

na poziomie państwa, regionów i instytucji. Badania prowadzone przez urzędy, które, mimo zastrzeżeń (niekompletność danych, gromadzenie zjawisk jedynie zarejestrowanych), w dużej części przyjmowane są jako dane prawdziwe i traktowane jako „obiektywne”. Dużą rolę odgrywa tu charakter urzędowy badań statystycznych, niepodlegający dyskusji poprzez fakt gromadzenia danych przez instytucje państwowe, w których obowiązują odpowiednie przepisy i sankcje za brak przesłania formularzy z danymi.

Statystyka oznacza dane liczbowe dotyczące agregacji danych z mniejszych jednostek do większych, ale też naukę o zbieraniu, analizie i interpretacji oraz prezentacji danych. Badania statystyczne są prowadzone na poziomie wojewódzkim, krajowym i międzynarodowym, stąd porównywalność wskaźników musi być zapewniana poprzez administracyjne wymogi i nadzór nad wypełnianiem formularzy lub metody reprezentatywne i standaryzacyjne porównywalności wskaźników (np. metodologia GUS, Eurostat). Badania statystyczne są podzielone na dziedziny społeczno-ekonomiczne: demografia, poziom i warunki życia, ubóstwo, opieka zdrowotna, zatrudnienie, *etc.* Podział ten nie jest zorientowany na realizację celów np. naukowych (weryfikacja teorii), lecz planistycznych i prognostycznych, w szczególności na zarządzanie różnymi dziedzinami, jak ludność, gospodarka czy inne zasoby.

W świecie publicznym pożądane są wyniki badań według pozytywistycznego czy też normatywnego paradygmatu i metodologii naukowej, gdzie dane są przedstawiane w postaci porównywalnych okresów, liczb, procentów, na wykresach i prezentacjach wizualnych. Badacze, pracownicy urzędów i mass mediów poszukują wyników obrazujących porównywalne zjawiska społeczne w czasie, a dane społeczne gromadzone przez urzędy stanowią zapisy stanów „obiektywnych”. W ten sposób są rejestrowane wydarzenia (urodzenia, zgony, zachorowania), zachowania (małżeństwa, rozwody, przestępstwa) i zasoby (zatrudnienie, wynagrodzenia, wykształcenie). Można też mówić o danych „subiektywnych” jako zapisach stanów świadomości, czyli deklaracjach narodowych lub preferencjach partyjnych na temat wyników wyborów parlamentarnych, prezydenckich lub samorządowych [Sułek 2002].

Badania statystyki urzędowej mają sens o tyle, o ile są badaniami porównań zbiorowości. Dzisiejsze badania statystyczne mają zupełnie inny charakter niż te, które były prowadzone w XIX wieku, gdy powstawały pierwsze koncepcje socjologiczne w tym obszarze lub nawet badania sprzed lat dwudziestu czy trzydziestu. Główne różnice leżą w unowocześnieniu systemów informacji, technologii i stałym powiększaniu oraz przepływie wiedzy [Oleński 2006]. Proces rejestracji i zbierania danych podlega zewnętrznej i wewnętrznej kontroli urzędowej.

Jeżeli chodzi o weryfikację formularzy statystycznych, to ze względu na wielkość zbioru w danym obszarze lub okresie, w statystyce urzędowej najczęściej sprawdza się jedynie niewielką wylosowaną podgrupę zdarzeń, w celu określenia zgodności rejestrowanych faktów z rzeczywistością. W badaniach statystycznych rola weryfikacji formularzy, polegająca na sprawdzeniu wpisów statystycznych u źródła, jest zbyt mała, aby można na tej podstawie zweryfikować całość badań.

Na terenie Unii Europejskiej w ramach statystyki społecznej są opracowywane w tym celu metody standaryzacji i porównywalności badanych zjawisk. Jeżeli idzie o badania z wykorzystaniem próby losowej, metody i techniki są takie same jak prowadzone przez ośrodki badań socjologicznych i opinii publicznej (por. wspólne elementy metodologii badań opinii społecznej i metodologii badań GUS, Eurostat; np. kwestionariusze ankiet komputerowych CATI i in. [Kordos 2007]). Ewentualne trudności dotyczą kwestii, z którymi spotykają się ankierzy w czasie wypełniania kwestionariuszy (np. niechęć do podawania niektórych danych osobowych lub finansowych).

Korzystanie z takich rozwiązań metodologicznych w badaniach statystycznych nie zwalnia z wątpliwości co do wiarygodności źródeł i rzetelności danych, gromadzonych przez urzędy. Chodzi o czynniki i uwarunkowania społeczne zachowań mających miejsce w organizacjach podczas zbierania danych [Sułek 2002]. Wpływ tych zachowań na uzyskiwane wskaźniki jest wciąż sprawą dyskusyjną wśród statystyków, braną pod uwagę w ustaleniach metodologicznych. Socjolog musi zwracać uwagę na pojawiające się problemy, i uwzględniać je przy definiowaniu badanych zjawisk i analizie danych już zebranych. W niektórych przypadkach problem może być trudny na poziomie rejestracji i publikacji danych. Między krajami występują różnice w zbieraniu tych samych danych, co przyczynia się do powstawania trudności w ujednoceniu definicji porównywalnych zjawisk [por. Taylor 1982].

Inna trudność dotyczy niekompletności danych i porównywalności okresów. Kompletność danych dotyczy przypadku, kiedy w badanym okresie są one publikowane ze wszystkich dostępnych źródeł. Przygotowanie badań społecznych z wykorzystaniem danych urzędowych opiera się na następujących elementach: (1) określenie definicji urzędowych, (2) kryteria definicyjne ustalone przez badacza, (3) zakres gromadzonych danych, (4) charakter rejestrowanych danych według warunków instytucjonalnych. Zainteresowania badaczy winny być zbieżne z obszarem sprawozdawczości urzędów.

Gromadzenie danych przez urzędy jest procesem, w którym uczestniczy się pracowników: od tych, którzy zbierają dane po takich, którzy je kodują, weryfikują, liczą i zestawiają. Etap pierwotny to ten, w którym dane są wytwarzane



i powstaje decyzja o treści wpisywanej do kwestionariusza ankiety (lub zostaje wstawiony kod oznaczający zjawisko). Gromadzenie danych z *wiarygodnych* źródeł zależy od wstawienia w kwestionariuszu odpowiedniej cyfry (kodu, znaku) lub pojęcia na etapie pierwotnym.

### WIARYGODNOŚĆ ŹRÓDEŁ URZĘDOWYCH

Ocena wiarygodności danych nie należy do zainteresowań badań, które są zorientowane na podejście konstrukcjonistyczne. Mimo to, zwłaszcza gdy mamy do czynienia z danymi publicznymi, warto wiedzieć, skąd biorą się „problemy” urzędów.

Według A. Sułka [2002] wiarygodność informacji zależy od różnych warunkowań społecznych wewnątrz organizacji zbierającej dane i w jej otoczeniu. Zaliczyć do nich można tendencyjne zachowania informatorów w instytucji, na które wpływ ma „chęć uzyskania (lub uniknięcia straty) pewnych dóbr oraz chęć pokazania się od dobrej strony, (...) zachowania są brane pod uwagę przy dystrybucji ważnych dla nich dóbr lub też są podstawą ich oceny” [Sułek 2002: 124].

Struktura instytucji ma wpływ na dane opisujące jej działalność i wyniki. Jeżeli istnieją konflikty i różnice wewnątrz organizacji, to ich elementem jest „dostarczanie jednych informacji a ograniczanie innych” [Sułek 1998: 143]. Podwładni skłonni są „zawyżać wskaźniki swej efektywności, wyolbrzymiać i ukrywać zasoby” [Sułek 1998, tamże]. Bardziej rzetelne są dane sprawozdawcze niezgodne z interesami instytucji niż ukazujące ją w dobrym świetle oraz przede wszystkim „dane zebrane przez jednostkę niezależną od instytucji, której działalność ukazują niż dane produkowane przez tę samą instytucję” [Sułek, tamże]. Skądinąd „dane zebrane przez instytucję na własny użytek będą bardziej rzetelnie niż te, które są przeznaczone dla instytucji zwierzchnich lub dla opinii publicznej” [Sułek, tamże].

W socjologii i naukach społecznych dane ilościowe są poddawane ocenie i weryfikacji metodologicznej, a ich wiarygodność jest oceniana krytycznie z punktu widzenia społecznych uwarunkowań rejestracji. Analizy interakcji w sytuacjach i kontekstach społecznych ujawniają ograniczenia gromadzenia danych urzędowych na przykładzie psychiatrii w celu zastosowania ich do wykorzystania w badaniach ilościowych [Garfinkel, Bittner 2007]. Jednocześnie zachęciło to socjologów do pogłębienia analiz jakościowych na ten temat.

W XIX stuleciu wprowadzono termin „statystyka moralna” na określenie zjawisk poruszanych przez opinię publiczną, a charakteryzujących się stałą czę-

stotliwością występowania, np. rozwodów, zapadalności na choroby śmiertelne i inne. Jednym z problemów społecznych były wspomniane już samobójstwa. Następnie badania socjologiczne z wykorzystaniem danych urzędowych prowadzili socjologowie różnych opcji, np. socjografowie jak np. Steinmetz [1913] czy Lazarsfeld, Jahoda i Zeisel [2007] w badaniu nad bezrobociem, godząc analizę jakościową ze statystyczną.

Jeżeli chodzi o ocenę wiarygodności informacji zbieranych przez urzędy statystyczne z kościołów, związków wyznaniowych i stowarzyszeń narodowościowych, należy ostrożnie z nich korzystać i je weryfikować. W Polsce kościoły nie są zobowiązane do składania obowiązkowych sprawozdań jak przed 1989 r. do Urzędu do spraw Wyznań, a jedynie na zasadzie dobrowolności przesyłają ankiety do urzędów statystycznych. Swoją statystyką i badaniami dysponuje Kościół katolicki. W corocznych badaniach są notowane (1) liczba wiernych w parafiach, (2) odsetek wiernych uczęszczających na niedzielą mszę w stosunku do ogólnej liczby ochrzczonych w parafii oraz (3) liczba wiernych, którzy przystępują do komunii św., odniesiona do liczby ochrzczonych<sup>2</sup>. Lucjan Adamczuk [2003] pisze o analizie ankiet: „Proste zestawienie liczb za ostatnie 10 lat wskazuje, że niektóre z wyznań podają dane, bez oparcia ich o stałe corocznie weryfikowane źródła ewidencyjne. Wskazuje to na brak ukształtowanych, stałych procedur zbierania informacji statystycznych, a być może i potrzeb, nawet na swój własny użytek.” [Adamczuk 2003: 9]. W terminologii badań statystycznych dotyczących wyznań znajdują się kwestie trudne i problemy pojęciowe, np. definicja liczby wiernych (w ankietach Kościoła katolickiego zastępuje liczba ochrzczonych), pojęcie duchownego czy też zróżnicowanie jednostki organizacyjnej na parafie, zbory i inne [Tamże].

### JAKOŚĆ DANYCH W OCHRONIE ZDROWIA

W obszarze nauk medycznych panuje przekonanie o potwierdzaniu koncepcji naukowych za pomocą badań przy wykorzystaniu metod ilościowych, najlepiej na dużych próbach reprezentatywnych. W dobrze przygotowanych badaniach medycznych inne metody obserwacji naukowej nie zdają egzaminu, i tym samym występuje zaufanie dla rejestrów lekarskich i danych szpitalnych gromadzonych metodami administracyjnymi, które mają pieczęć instytucji państwowych. Statystycy medyczni przestrzegają jednocześnie o istnieniu wiarygodnych i „rzetelnych źródeł” w szpitalach, które mogą być zawsze lepsze jakościowo

<sup>2</sup> Źródło: Instytut Statystyki Kościoła Katolickiego SAC. Strona WWW.

tylko dzięki uświadamianiu w tym kierunku personelu medycznego [Wojtyniak, Goryński 2006]. Socjologowie przypominają, że rzetelność informacji odnosi się „do stopnia zgodności, w którym przypadki są przypisywane tej samej kategorii przez wielu obserwatorów lub przez jednego obserwatora w różnorodnych okolicznościach” [Hammersley 1992, za: Silverman 2008: 245]. Rzetelność jest cechą procedur pomiarowych, której zastosowanie polega na wielokrotnym uzyskaniu tego samego lub zbliżonego wyniku [Konecki 2000].

Proces „konstruowanie społeczne” rozpoczyna się już w momencie opisu przypadku choroby psychicznej, kiedy pacjent poda inne objawy, albo lekarz – gdy błędnie wpisuje inny kod choroby. Gdy w procesie badawczym nie zaszedł żaden błąd na wcześniejszym etapie, to badacz tendencyjnie może interpretować dane, np. będąc zwolennikiem poglądu, że wśród białych Amerykanów jest mniej chorych psychicznie niż wśród innych ras i narodowości zamieszkujących te same tereny [Bastide 1972].

Na przykład odnośnie dużych porównawczych badań międzynarodowych, wydatki służby zdrowia są mierzone za pomocą różnych wskaźników i podawane w rankingach gospodarczych (OECD, Eurostat). Opłacalność działań szpitala można mierzyć za pomocą liczby leczonych pacjentów, analiz wskaźników chorobowości w danym regionie oraz badania zadowolenia z opieki medycznej. Wobec dyskursu publicznego na temat służby zdrowia na innym poziomie pozostają dociekania metodologiczne, według których można ustalić rzeczywistą wartość danych.

Do obowiązków osób wypełniających formularze statystyczne należy uczestnictwo w okresowych szkoleniach organizowanych przez szpital. Rejestracja danych bywa uciążliwym i nieprzyjemnym obowiązkiem administracyjnym, dlatego w niektórych krajach pracownicy szpitali są dodatkowo wynagradzani za wypełnienie formularzy statystycznych. W różnych dziedzinach statystyki medycznej są prowadzone badania weryfikujące etap rejestracji i zbierania danych pierwotnych, to jest jakości procesu mającego wpływ na ocenę wiarygodności źródeł. Lekarz Adam Koziarkiewicz [1999] na podstawie badania statystyki chorobowości szpitalnej ustalił następujące cechy procesu kodowania kart chorobowych: (1) niekompletność danych, ponieważ w badanym okresie tylko część jednostek została zarejestrowana według odpowiednich kategorii chorobowych, (2) brak szczegółowości zapisu, którą można było ustalić na podstawie porównania liczby przypadków chorobowych zarejestrowanych przez pracowników szpitala i liczby przypadków kontrolnych, wpisanych przez badacza według obowiązującej klasyfikacji chorób, (3) niespójność rejestracji kategorizowanych

przypadków według klasyfikacji: okazało się, że dużą część chorób przyporządkowano do niewłaściwych kategorii statystycznych.

Wnioskiem podsumowującym badanie jest niekompletność danych na poziomie instytucji sprawozdającej (tj. oddziału szpitalnego) z powodu problemów organizacyjnych i kadrowych szpitala, a także z powodu braku szkoleń w tym zakresie. Innym uwarunkowaniem jakości zapisu i rejestracji kart jest niechęć do wypełniania obowiązków statystycznych przez lekarzy i przekazywanie ich innym osobom, które zbyt szeroko definiują i interpretują zapisywane zjawiska [Kozierkiewicz 1999; Kozierkiewicz, Sala 1997]. W nieco podobny sposób analiz dokonują etnostatystycy, ale nie sprawdzają wiarygodności dokumentów, lecz badają ich sens kulturowy (dalej).

### REJESTRACJA DANYCH A ZNACZENIE ZJAWISKA. ABORCJA

Przykładem konstruowania społecznego danych urzędowych są zabiegi aborcyjne. W Polsce obowiązuje ustawa dopuszczająca wykonywanie przerywania ciąży w trzech przypadkach: z powodu ciąży zagrażającej życiu lub zdrowiu matki, w wypadku wady wrodzonej płodu oraz w wypadku czynu zabronionego (gwałtu). Jednak co roku w szpitalach rejestruje się nieliczne dane urzędowe: w 2009 r. zarejestrowano 538 przerwania ciąży, a policja informuje o około 100 zgłoszeniach na temat zabiegów dokonanych poza prawem [*Sprawozdanie z wykonania Ustawy 2010*]. Ponadto istnieje ciemna liczba zjawiska. Nieliczne dane to skutek prawnej i finansowej zależności instytucji medycznych od zwierzchników politycznych, którzy tak zapewniają szpitalom określone środki na funkcjonowanie i leczenie. Otrzymanie skierowania od lekarza specjalisty na zabieg lub badanie genetyczne jest niezwykle trudne do realizacji, ponieważ nawet w przypadku okazania takiego skierowania inni lekarze odmawiają wykonania aborcji korzystając z prawa do klauzuli sumienia. Często jest to też wynik zakazów wydawanych przez kierowników i ordynatorów, którzy nie chcą, aby w statystyce figurował zabieg aborcji; dlatego zabiegi realizowane są drogami nieformalnymi. Dane o przerywaniu ciąży są publikowane w sprawozdaniu rządowym, stanowiącym przedmiot badania na temat konstruowania społecznego, i zawierającym przede wszystkim dane wytworzone i opisane przez szpitale i urzędy, [Maciąg 2007].

Przed zmianą prawa aborcyjnego księża katoliccy w Polsce głosili dokonywanie aborcji do 1 miliona, gdy obliczenia szacunkowe wynosiły 620–700 tysięcy, a dane urzędowe – ok. 165 tys. [Okólski, 1988]. Te i inne działania spowodowały zaokrąglenie liczb, będące na tyle skuteczne, że doprowadziło to do

zakazu prawnego aborcji w 1993 r. Według obliczeń K. Meissnera [1989] w czasie obowiązywania ustawy dopuszczającej bezpłatną aborcję w szpitalach dokonano 21 milionów zabiegów. W czasie największej dyskusji na temat wprowadzenia zakazu aborcji na początku lat 90. rejestrowano ok. 80 tysięcy zabiegów, a liczba ta systematycznie malała [źródło: *Rocznik Statystyczny Ochrony Zdrowia* 1997].

Dlatego nie wiadomo, ile rzeczywiście dokonuje się aborcji i w tej sytuacji w sporach społecznych o zmianę prawa dane na temat zabiegów są przedstawiane odmiennie przez działaczy antyaborcyjnych i ruchy kobiece. obrońcy życia dzieci nienarodzonych szacują, że corocznie wykonuje się ok. 7–14 tys. przerwania ciąży [Stachura i In. 2003]. Ruchy kobiece podają dane w paśmie 80–240 tys. [Nowicka 2007]. Nie wiadomo też, w jaki sposób obliczono te wielkości, chociaż organizacje te podają jako źródła przede wszystkim argumenty demograficzne (np. ruchy kobiece podają spadek liczby urodzeń jako przyczynę aborcji, a obrońcy życia powołują się na metody stosowane przez demografów w latach 80.). Innego zdania jest Igor Jaruga [1999], który przeprowadził badanie z uwzględnieniem metodologii porównawczej dla dwóch okresów w latach 90., i oparł je na wskaźnikach dotyczących płodności kobiet i korzystania ze środków antykoncepcyjnych: wskaźnik wyniósł ok. 129–100 tys. zabiegów o tendencji spadkowej (przyczyną jest zapobieganie ciąży za pośrednictwem antykoncepcji), [por. Maciąg 2007].

Joel Best [2001] wyróżnia zjawisko wyolbrzymiania danych (lub pomniejszania), a także mechanizm zaokrąglania liczb dla uzyskania lepszego efektu w przekazie społecznym. Jeśli chodzi o metodologię liczenia danych, występuje zjawisko błędnego obliczania i braku umiejętności liczenia, a także podawanie wyników według „własnej metodologii”, sprowadzającej się do postępowania według kryteriów, które bardziej pasują ideologom organizacji, ruchów społecznych lub partii politycznych, dane te publikujących.

Częstym argumentem są dane przedstawiane w kontekście przekonywania grup społecznych, mających odmiennie cele i poglądy; wykorzystują je autorytety i eksperci – politycy, lekarze, księża. W tym przypadku z reguły nie dba się o wiarygodność informacji, ale o wymiar społeczno-polityczny przekazu medialnego oraz o to, „kto mówi” i jakie wartości reprezentuje. W omawianiu problemów społecznych dane są argumentami politycznymi.

Jednostki definiują świat społeczny za pomocą wspólnych znaczeń języka (metafory, wielkie liczby), których kategorie są wyolbrzymiane w celu uzyskania specjalnego efektu. Jest to cecha specyficzna i naturalna komunikacji społecznej, w której są konstruowane znaczenia w sytuacjach i kontekstach społecznych. W rozmowie potocznej dane są przytaczane z pamięci i najczęściej w zaokrągleniu, dlatego zjawiska społeczne są wytworami sytuacji i wyobrażeń społecznych.

Źródeł nielicznych danych urzędowych można dopatrywać się w kilku przyczynach. W szpitalach rejestruje się: (1) przerwania ciąży (z powodów wspomnianych ustawowo) oraz (2) poronienia samoistne. Określenie definicji urzędowej przerwania ciąży nie jest jasne, ponieważ wśród zarejestrowanych poronień samoistnych [ok. 40 tys. rocznie; *Sprawozdanie z wykonania Ustawy 2010*] występują przypadki, w których zastosowano wcześniej aborcję farmakologiczną [Maciąg 2007]. Zapis w dokumentacji szpitalnej jest raczej odzwierciedleniem stanu prawnego a nie faktycznego, nawet gdy zastosowano środek farmaceutyczny. Dlatego kwestię tę można traktować jako konstrukcję społeczną, i analogiczną do problemu wskaźników samobójstw, ponieważ ustalenie stanu zdrowotnego w dokumentacji jest możliwe po uwzględnieniu informacji, że ciąża nie przebiega typowo z punktu widzenia okresowych badań medycznych w historii ciąży [z wywiadów z lekarzami i położnymi, Maciąg 2007]. Zapis lekarski wygląda różnie: termin „aborcja” pochodzi od łacińskiego terminu „abortio”, oznaczającego poronienie płodu [Encyklopedia PWN 2010], lecz jest pomijany przez ginekologów, a w języku fachowym funkcjonuje ogólny termin „przerwanie ciąży” lub „poronienie” oraz kilkanaście pojęć szczegółowych, np. poronienie zatrzymane, tj. *łacińskie* *abrazja*, poronienie w toku *etc.* [Komender 1996].

### TEKSTY, DOKUMENTY I AKTA. ETNOSTATYSTYKA

Dokumenty i dane urzędowe są badane jako teksty wytwarzane w organizacjach [Silverman 2008]. Dlatego konstruowanie społeczne odnosi się również do analizy dokumentów, czyli zainteresowania różnymi rodzajami pism wytwarzanych przez instytucje. Według Paula Atkinsona i Amandy Coffey [1997] analiza treści dokumentów przeprowadzana jest w celu poznania tekstów wytwarzanych w czasie codziennej pracy. Odbiorcy komunikatów zamieszczonych w dokumentach są przekonywani do obiektywności prezentowanych treści w celu oceny danej instytucji za pomocą literackich środków wyrazu i opisów językowych. Socjolog zwraca uwagę na komentarze do danych, tabel lub podliczeń finansowych, interesujących go ze względu na „organizację społeczną” dokumentów. W ten sposób może dowiedzieć się o charakterze planowanych przedsięwzięć, historii i rodzaju pracy, mając na uwadze funkcjonowanie instytucji, np. przeprowadzając analizę raportu uniwersyteckiego [Atkinson, Coffey 1997: 45 i n., tamże]. Praca instytucji jest uzależniona od prowadzenia dokumentacji spraw, czyli od pamięci zbiorowej, zapisywania, przechowywania, archiwizowania i ciągłego odtwarzania zasobów urzędowych, coraz częściej w formie elektronicznej.

Perspektywę konstrukcji i organizacji społecznej dokumentów i danych można nazwać inaczej społeczną koncepcją powstawania i kreowania danych urzędowych przez pracowników instytucji. Chodzi o opis zbierania i przetwarzania informacji pisanej. Badanie i interpretacja dokumentów w miejscu pracy i w organizacjach wytwarzających teksty w ogóle, jest prowadzona w zakresie różnych dziedzin (np. raporty kościelne, medyczne, policyjne lub sądowe). Badanie odkrywania natury znaczeń społecznych w rzeczywistości instytucjonalnej ma sens pod warunkiem wyjaśniania pojęciowego konkretnych kategorii danych. Badania statystyczne i komentarze do nich poddaje się interpretacji jako teksty pod kątem weryfikacji językowej (analiza etnograficzna, kulturowa) i weryfikacji wniosków ilościowych.

W celu otrzymania pełniejszego obrazu powstawania danych, w najnowszych badaniach etnograficznych w organizacji proponuje się śledzenie udostępniania danych, upoważnień i sygnatur, logowania w sieci komputerowej poświadczonych odpowiednimi aktami lub wyciągiem z e-maili. W tym nurcie analizy kulturowej Robert P. Gephart [1988] zaproponował badanie etnostatystyki w ośrodku przygotowującym raporty statystyczne, polegające na obserwacji wytwarzania, interpretacji i prezentacji danych ilościowych [Gephart 1988: 9 i n.]. Przedrostek *'ethno-'* oznacza aspekty kulturowe w analizie socjologicznej, w której wykorzystywane są też pojęcia etnometodologii w pracy nad tworzeniem danych, interpretacją i pisaniem raportów końcowych [Gephart 2003]. Badanie etnostatystyczne polega zatem na analizie językowej i retorycznej tekstu oraz na wyszukiwaniu szczegółowych zastosowań w treści komentarzy do danych urzędowych. Takie badanie dokumentów polega na analizie treści przekazów masowych i tekstów literackich, stanowiąc niejako przeciwieństwo do analiz skrótowych opisów i tabel danych statystycznych, w których obowiązują zasady i techniki liczenia, agregacji i szacowania danych.

R. Gephart wyróżnia trzy obszary badań etnostatystycznych: (1) badanie na temat produkcji statystyki (*studying the production of statistics*), (2) zestawienia w działaniu (*statistics at work*), (3) statystyka pod kątem retoryki (*statistics as rhetoric*) [por. tamże Gephart 1988: 9–10]. W pierwszym z wymienionych aspektów prowadzenia badań nad rutynowym procesem powstawania danych w organizacjach, najważniejszym elementem jest kultura organizacyjna pracy badaczy, to jest opis zależności i hierarchii w instytucji, w której dane powstają. Mniej interesujący jest aspekt ilościowy statystyk.

Drugi obszar badania etnostatystycznego, obejmujący statystyki w działaniu, czyli sprawdzenie założeń technicznych (istotność korelacji, wnioski z badań). Produkcja danych dotyczy zjawisk mierzalnych, to jest szacowania różnych

parametrów, eksperymentów i modeli przyczynowych, dlatego dane pierwotne zostają dokładnie zweryfikowane. Obserwacja dotyczy również zestawień i interpretacji tekstów i komentarzy do tabel oraz sprawdzeniu, na ile są one przygotowane i wyłożone dla uniwersalnego odbiorcy, w konwencji przyjętej w danej instytucji [Gephart 1988: 29-46].

Trzeci poziom etnostatystyki obejmuje dane i teksty w analizie retorycznej. Taka analiza tekstu przydatna jest w różnych rodzajach badań publicznych. Badanie naukowe mają na celu informowanie niejako *poprzez* teksty, to jest opisy i komentarze pod tabelami statystycznymi i wykresami. I w tej perspektywie raporty publiczne traktowane są jako wydarzenia *per se*, czyli są analizowane jako teksty z punktu widzenia praktycznych form literackich. Przykładem jest analiza wyliczeń finansowych, w której badacz wykorzystuje krytyczną analizę literacką, a znaczenia treściowe umieszcza w kategoriach społecznych zgodnych z różnymi pojęciami i regułami zastanymi w organizacji [Gephart 1988: 47-61].

Język prezentacji raportów publicznych dotyczy określonego „centrum uwagi”, tzn. właściwego nakierowania jej na odbiorcę, i poszukuje „literackich, symbolicznych i lingwistycznych technik i metod, poprzez które realizowane jest obrazowanie perswazyjne; metody projektujące wywoływanie przekonań” [Gephart 1988: 28]; są to takie narzędzia jak pogrubienie czcionki, kolorowy wykres, czerwone podkreślenie, itd. Owo „centrum uwagi” nie służy wyrażaniu danych statystycznych w zwykłej retoryce lub środkach literackich, lecz jest raczej perspektywą potwierdzającą ważność dokumentów naukowych. Badania etnostatystyczne mogą mieć praktyczne implikacje dla socjologa, mającego możliwość korzystania z danych statystycznych i opisów, zwłaszcza z danych i zmiennych ilościowych, który zastosuje je dla interpretacji jakościowej.

## PODSUMOWANIE

Znaczenie danych urzędowych zależy od społecznej definicji sytuacji i konstruowania społecznego. Zdaniem J.D. Douglasa [1967] fakt społeczny zapisany w statystyce jest na tyle obiektywny, na ile jego znaczenie społeczne zostało wynegocjowane i nazwane w danej sytuacji społecznej. Dlatego według etnometodologów cele badania winny stanowić obserwację pierwotnego procesu gromadzenia danych [np. Garfinkel, Bittner 2007]. Dzisiaj analizy w tym zakresie przeprowadzają etnografowie, również w instytucjach medycznych [Hammerley, Atkinson 2000]. Wytwarzanie danych społecznych ma bowiem związek z definiowaniem sytuacji w toku interakcji podczas zapisywania informacji



źródłowej oraz z ustalaniem klasyfikacji zachowań społecznych pracowników urzędów [Cicourel 1968].

Konstrukcja społeczna oznacza statystykę lub dokument urzędowy wytworzony i publikowany przez organizację (ministerstwo, bank, korporację zawodową *etc.*). Analiza dokumentów jest możliwa we współpracy z daną instytucją, gdy socjolog staje się obserwatorem uczestniczącym zachowań i procesu tworzenia dokumentów oraz ma dostęp do akt lub archiwów. Takie postępowanie jest możliwe w badaniu niektórych zachowań i sytuacji społecznych w organizacji, gdy badacz ma dostęp do danych lub materiałów albo sam jest w nie w jakimś stopniu zaangażowany. Np. socjolog może uczestniczyć w niektórych czynnościach urzędu lub korporacji.

Dane urzędowe obrazują zmiany zachodzące w społeczeństwie i gospodarce. Badania wskaźników samobójstw i krytyczne podejście do koncepcji konstruowania społecznego przetrwały się z czasem w pogłębione studia nad przydatnością danych w urzędach statystycznych i medycznych [np. Taylor 1982]. Jednakże warto zadać pytanie, dlaczego badacze rzadko korzystają z zasobów administracyjnych w celu analiz socjologicznych (np. ze spisów powszechnych, rejestrów ludnościowych itd.). Według Martina Bulmera [1980] do niewielkiego zainteresowania statystyką urzędową wśród socjologów przyczynił się niewielki zakres wykorzystania wśród propozycji badań i danych urzędowych, a zaufanie do danych jest przedstawiane w sposób wątpliwy przez samych socjologów. Bulmer zauważa, że socjologowie nie korzystają z danych urzędowych z powodu małej wiarygodności źródeł i nierzetelnych danych. Jego zdaniem nie ma potrzeby adaptacji danych urzędowych do badań nauk społecznych, ponieważ są one jedynie ubocznym produktem działalności organizacyjnej i administracyjnej urzędów. Przyczyn trudności i niechęci do wykorzystywania danych urzędowych należy szukać w innych celach badań społecznych i badań statystycznych oraz w innym zdefiniowaniu badanych zjawisk. Krytyka danych urzędowych jest rozwijana z różnych punktów widzenia na problemy społeczne, m.in. konstrukcjonistycznego, w których idzie o wąskie spojrzenie jedynie w kilku dziedzinach (samobójstwa, dewiacje), a praktycznie brak w niej rozwinięcia innych analiz o charakterze społecznym, mogących stać się podstawą do szerszej dyskusji na temat wykorzystania danych urzędowych w socjologii. Wypada dodać, że Bulmer przedstawiał swoje stanowisko w chwili, kiedy w socjologii zachodniej dochodziło do podsumowań fali badań wzorujących się na koncepcjach interakcjonistycznych i fenomenologicznych, a miał nastąpić dopiero rozwój analiz dokumentów w stronę etnografii organizacji i badań języka [Silverman 2008; Lincoln, Denzin, 2009]. Dlatego na podstawie obserwacji w organizacjach re-

alizowane są badania etnograficzne, i tym samym rzeczywistość instytucjonalną dobrze jest analizować poprzez język dokumentów i retorykę tekstu.

Społeczne uwarunkowania gromadzenia danych, np. w sprawie aborcji lub bezrobocia, mają związek z nadawaniem społecznych znaczeń w instytucjach. Liczby są często elementem dyskursu społecznego ze względu na uczestnictwo zainteresowanych grup społecznych i wykorzystanie języka w aktualnym problemie społecznym, angażującym społeczeństwo. W przestrzeni społecznej i medialnej jest to interpretacja podanej liczby i zabieg definiowania zjawiska poprzez ustosunkowanie się do problemu (bezrobocie, aborcja). W wyniku podania najnowszych danych zarejestrowanych przez urzędy nie tylko wypowiadają się politycy, lecz sytuację definiują też odbiorcy, czyli politycy, dziennikarze, naukowcy, eksperci, i sami zainteresowani, np. bezrobotni, chorzy na raka, feministki. W tym sensie statystyki można badać jako wytwory sytuacji i relacji społecznych, które mają różne znaczenia dla jednostek, grup i instytucji społecznych.

## BIBLIOGRAFIA

- Adamczuk L. [2003]: *Statystyka wyznań religijnych w Polsce*, [w:] L. Adamczuk, G. Gudaszewski (red.), *Wyznania religijne. Stowarzyszenia narodowościowe i etniczne w Polsce 2000-2002*, Warszawa: Główny Urząd Statystyczny.
- Atkinson P., Coffey A. [1997]: *Analysing documentary realities*, [w:] D. Silverman (red.), *Qualitative Research: Theory, Method and Practice*, London: Sage Publications.
- Bastide R. [1972]: *Socjologia chorób psychicznych*, Warszawa: Państwowe Wydawnictwo Naukowe.
- Becker H.S. [2009]: *Outsiderzy. Studia z socjologii dewiacji*, Warszawa: Wydawnictwo Naukowe PWN.
- Berger P., Luckmann T. [1983]: *Społeczne tworzenie rzeczywistości*, Warszawa: Państwowy Instytut Wydawniczy.
- Best J. [2001]: *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*, Berkeley and Los Angeles: University of California Press.
- Blumer H. [2007]: *Interakcjonizm symboliczny. Perspektywa i metoda*, Kraków: Zakład Wydawniczy „Nomos”.
- Bulmer M. [1980]: *Why don't sociologists make more use of official statistics?*, “Sociology”, 4(14).
- Cicourel A.V. [1964]: *Method and Measurement in Sociology*, New York: Free Press of Glencoe.
- Douglas J.D. [1967]: *The Social Meanings of Suicide*, New Jersey: Princeton University Press.
- Durkheim É. [2006]: *Samobójstwo. Studium z socjologii*, Warszawa: Oficyna Naukowa.
- Frysztański K. [2009]: *Socjologia problemów społecznych*, Warszawa: Wydawnictwo Naukowe „Scholar”.
- Garfinkel H., Bittner E. [2007]: *Dlaczego kartoteki pacjentów muszą być prowadzone “źle”?*, (w:) H. Garfinkel, *Studia z etnometodologii*, Warszawa: Wydawnictwo Naukowe PWN.

- Glaser B.G., Strauss A.L. [2009]: *Odkrywanie teorii ugruntowanej. Strategie badania jakościowego*, Kraków: Zakład Wydawniczy „Nomos”.
- Gephart R.P. [1988]: *Ethnostatistics. Qualitative Foundations for Quantative Research*, London: Sage Publications.
- Gephart R.P. [2003]: *Ethnostatistics* (Hasło), [w:] *The Encyclopedia Social Science Research Methods*, Vol. 1., M.S. Lewis-Beck, A. Bryman, T. Futing Liao, London: Sage Publications.
- Halbwachs M. [1930]: *Les Causes du Suicide*, Paris: Alcan.
- Hammersley M. [1992]: *What's Wrong with Ethnography: Methodological Explorations*, London: Routledge.
- Hammersley M., Atkinson P. [2000]: *Metody badań terenowych*, Poznań: Zysk i S-ka Wydawnictwo.
- Jarosz M. [2004]: *Samobójstwa. Ucieczka przegranych*, Warszawa: Wydawnictwo Naukowe PWN.
- Jaruga I. [1999]: *Bezpośrednie determinanty płodności. Aplikacja modelu Bongaartsa dla Polski dla lat 1991 i 1995*, „Studia Demograficzne” 1999 z. 2.
- Karpiński J. [2006]: *Badania diachroniczne*, Tenże: *Wprowadzenie do metodologii nauk społecznych*, Warszawa: Wydawnictwo WSPiZ im. L. Koźmińskiego.
- Kitsuse J., Cicourel A.V. [1963]: A Note on the Uses of Official Statistics, “Social Problems”, 1963 z. 11.
- Komender J. (red.) [1996]: *Wielki słownik medyczny*, aut. hasel Z.J. Brzeziński i In.; PAN, Wydz. N. Med., Warszawa: Wydawnictwo Lekarskie PZWL.
- Konecki K. [2000]: *Studia z metodologii badań jakościowych. Teoria ugruntowana*, Warszawa: Wydawnictwo Naukowe PWN.
- Kordos J. [2007]: *Statystyczne badanie zjawisk społecznych*, [w:] T. Panek (red.). *Statystyka społeczna*, Warszawa: PWE.
- Koseła K., Sułek A. [2005]: *Badania wzdłuż czasu*, [w:] W. Kwaśniewicz (red.), *Encyklopedia socjologii*. Suplement, Warszawa: Oficyna Naukowa.
- Kozierkiewicz A. [1999]: *Chorobowość szpitalna – jakość kodowania*, „Zdrowie Publiczne” 1999 z.2.
- Kozierkiewicz A., Sala D. [1996]: *Karta statystyczna do karty zgonu*, „Antidotum – Zarządzanie w opiece zdrowotnej” 1996 z. 11–12.
- Lazarsfeld P. [2007]: *Przedmowa do nowego wydania*, [w:] M. Jahoda, P.F. Lazarsfeld, H. Zeisel (red.), *Bezrobotni Marienthalu*, Warszawa: Oficyna Naukowa.
- Maciąg R. [2007]: *Spoleczna konstrukcja i społeczne wykorzystanie danych medycznych. Przypadek aborcji*. Niepubl. praca doktorska, Instytut Socjologii UW.
- Meissner K.W. [1991]: *Częstość poronień*, „Słowo powszechne”, Nr 85-86(XLV), 12-13-14.04.1991.
- Nowicka W. (red.), [2007]: *Prawa reprodukcyjne w Polsce. Skutki ustawy antyaborcyjnej*, Warszawa: Federacja na Rzecz Kobiet i Planowania Rodziny, <http://www.federa.org.pl/?page=static&static=5&lang=1> pobrane styczeń 2010
- Okólski M. [1988]: *Reprodukcja ludności a modernizacja społeczeństwa: polski syndrom*, Warszawa: Książka i Wiedza.
- Oleński J. [2006]: *Infrastruktura informacyjna państwa w globalnej gospodarce*, Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

- Pescosolido B.A., Mendelsohn R. [1986]: *Social Causation or Social Construction of Suicide? An Investigation into the Social Organization of official Rates*, "American Sociological Review", T. 51, February.
- „Przerywanie ciąży” (hasło), [w:] Encyklopedia PWN, <http://encyklopedia.pwn.pl/> pobrane styczeń 2010.
- Rocznik Statystyczny Ochrony Zdrowia 1997. Warszawa: Główny Urząd Statystyczny 1998.
- Schütz A. [2006]: Potoczna i naukowa interpretacja ludzkiego działania, [w:] A. Jasińska-Kania, L.M. Nijakowski, J. Szacki, M. Ziółkowski (red.), *Współczesne teorie socjologiczne*, T. 2, Warszawa: Wydawnictwo Naukowe „Scholar”.
- Spector J.I., Kitsuse M. [1977]: *Constructing Social Problems*, Benjamin-Cummings Publ. Company.
- Silverman D. [2008]: *Interpretacja danych jakościowych. Metody analizy rozmowy, tekstu i interakcji*, Warszawa: Wydawnictwo Naukowe PWN.
- Stachura M., Trzcńska B., Wronicz J., Urban K. [2003]: *Podziemie aborcyjne – mity i fakty*, maszynopis, Kraków: Instytut Naturalnego Planowania Rodziny.
- Sprawozdanie Rady Ministrów z wykonania w roku 2010 Ustawy z dnia 7 stycznia 1993 r. o planowaniu rodziny, ochronie płodu ludzkiego i warunkach dopuszczalności przerywania ciąży (Dz.U.1993, Nr 17, poz.78 z późn. zm.).
- Steinmetz R.T. [1913]: *Die Soziographie in der Reihe der Gesellschaftswissenschaft*, „Archiv für Rechts- und Wirtschaftsphilosophie”, T. 6.
- Sułek A. [1998]: *Dokumenty urzędowe*, (w:) W. Kwaśniewicz (red.), Encyklopedia Socjologii, T. 1, Warszawa: Oficyna naukowa.
- Sułek A. [2002]: *Wiarygodność źródeł i rzetelność danych urzędowych*, [w:] Tenże, Ogród metodologii socjologicznej, Warszawa: Wydawnictwo Naukowe „Scholar”.
- Taylor S. [1982]: *Durkheim and the Study of Suicide*, London: The Macmillian Press.
- Weber M. [2002]: *Gospodarka i społeczeństwo. Zarys socjologii rozumiejącej*, Warszawa: Wydawnictwo Naukowe PWN.
- Wojtyniak B., Goryński P. [2006]: *Opis Systemu Informacyjnego w Ochronie Zdrowia w Polsce. Elementy oceny*, Warszawa: Państwowy Zakład Higieny.

Rafał Maciąg

#### THE SOCIAL CONSTRUCTION OF OFFICIAL DOCUMENTS AND DATA. MEDICAL DATA

##### Abstract

Social construction means that the document or data are a product of the organization in which they arise. The official document is so reliable, what the organization wants to show their achievements and to hide flaws. In this way are product official reports, church or bank. Special types of medical data are: suicide, psychiatric and abortion data. The article discusses the problem of low confidence sociologists in the official data and greater acceptance for the administration of quantitative data in medical. It is noted that the phenomenological and ethnomethodological concepts they contributed to the development of qualitative analysis in sociology. A specific example is the registration of suicide and the definition of situation analysis the primary data.

Official nature of the data and documents as a product of the institutions associated with the separate treatment of the social process, which is the stage of data collection and preparation and publication. The concept of reliability of the data by researchers adopted is to understanding the concept of science, since – in terms of phenomenological-constructive – documents and data are considered *per se*, as study are an independent research subjects. Such research process can be analyzed the data at the time of formation, that is, at the time of collection in the organization.

**Key words:** medical data, official documents, qualitative analysis.

JOLANTA LISEK-MICHALSKA  
MARCIN LIPIEC  
ANNA OLCZYK  
Uniwersytet Łódzki

## OCENA METODOLOGICZNEJ I PRAKTYCZNEJ JAKOŚCI FOKUSA – PROPOZYCJA PROCEDURY

### Streszczenie

Naczelnym celem praktyków badań FGI jest pozyskanie informacji umożliwiających zleceniodawcom badań rozstrzygnięcie pytań problemowych lub problemów decyzyjnych – uzyskanie informacji różnorodnych, nasyconych i unikalnych. Na podstawie naszych badań stwierdzamy, że na jakość merytoryczną badania FGI wpływa siła aktywacji poszczególnych efektów grupowych. Precyzując, wymiar nasycenia informacji jest regulowany przez siłę aktywacji efektu stymulacji i śnieżnej kuli, z kolei dwa pozostałe wymiary (różnorodność i unikalność informacji) są pochodną stopnia nasilenia efektu bezpieczeństwa, spontaniczności oraz syndromu grupowego myślenia. Skoro istnieje przełożenie między metodologiczną a praktyczną jakością badania FGI, warto poszukiwać czynników mogących różnicować siłę aktywacji efektów grupowych. Nasz artykuł jest propozycją uwzględniającą wskaźniki i zasady pozwalające w sposób zobjektywizowany sprawdzić, czy analizowana zmienna (np. cecha osobowości dyskutanta) jest czynnikiem faktycznie różnicującym nasilenie efektów grupowych i tym samym pośrednio oddziałującym na praktyczną jakość badania FGI.

**Słowa kluczowe:** fokus group interview, dynamika grupowa, jakość badań fokusowych

---

\* jolanta\_lisek@wp.pl

## WPROWADZENIE

Celem artykułu jest przedstawienie propozycji sposobu oceniania metodologicznej i praktycznej (merytorycznej) jakości badań fokusowych. Przyjmując za punkt wyjścia naszych rozważań szczególną cechą tego typu badania, mianowicie fakt, że uzyskiwany przez badacza materiał jest produktem grupy, a nie sumą indywidualnych wypowiedzi uczestników, uznaliśmy, że właśnie od sposobu, w jaki funkcjonuje w trakcie sesji grupa fokusowa zależy wartość zgromadzonych informacji. Naturalnie mamy świadomość że to, co zachodzi w trakcie sesji w badanych i pomiędzy nimi jest uzależnione także od innych czynników, chociażby takich jak zachowanie moderatora, scenariusz, warunki, w jakich odbywa się badanie, znaczenie, jakie dyskutowanym problemom przypisują badani i jeszcze wielu innych. Jednak w naszym projekcie akcent został położony właśnie na grupę i zachodzące w niej relacje i ich wpływ na jakość uzyskanego materiału. Dlatego też pod pojęciem „metodologiczna” jakość fokusa rozumiemy tutaj stopień nasilenia (in. siłę aktywacji) efektów grupowych, a jakość „praktyczna” (in. merytoryczna) to stopień bogactwa informacji uzyskanych z wykorzystaniem tej metody, określane na wymiarze różnorodności, unikalności i nasycenia.

Próba opracowania autorskiej propozycji oceny jakości badania FGI (fokus group interview) została podyktowana głównie brakiem publikacji w zakresie tej problematyki przy stałej eksploatacji badań fokusowych zarówno w obszarze marketingu jak i badań społecznych. Tym samym naszym celem jest uzupełnienie literatury koncentrującej się dotychczas na przedstawieniu sposobu realizacji omawianej metody o refleksję nad jakością fokusa. W końcowej części artykułu zaprezentujemy fragmenty wyników badań własnych przeprowadzonych techniką eksperymentu naturalnego, relacjonujące zależność między siłą aktywacyjną poszczególnych efektów grupowych (czyli jakością metodologiczną fokusa), a wymiarami praktycznej (merytorycznej) jakości fokusa.

Schemat postępowania badawczego przedstawiał się następująco:

Zadanie pierwsze polegało na ustaleniu, czy i w jaki sposób profil osobowości dyskutanta ujmowany na wymiarze konformizm – nonkonformizm, diagnozowany na podstawie wyników Kwestionariusza KANH-III różnicuje następujące wymiary funkcjonowania badanego w moderowanej dyskusji: 1) częstość werbalizacji wypowiedzi opozycyjnych (odmiennych od opinii poprzedników), 2) częstość werbalnych reakcji potakujących, 3) częstość artykulacji wypowiedzi stymulujących grupę do śmiechu, 4) częstość wtrąceń (wejść w słowo), 5) liczbę skojarzeń projekcyjnych wygenerowanych przez dyskutanta na potrzeby techniki chińskiego portretu, 6) częstość bezpośrednich nawiązań do wypowiedzi roz-

mówców, a także 7) częstość kierowania pytań do interlokutorów oraz 8) łączną liczbę wypowiedzi wygenerowanych przez dyskutanta.

Zadanie drugie miało umożliwić sprawdzenie, czy profil osobowości dyskutantów różnicuje stopień nasilenia wybranych efektów grupowych: efektu stymulacji, śnieżnej kuli, spontaniczności, bezpieczeństwa, grupowego myślenia i konformizmu normatywnego.

I wreszcie (zadanie trzecie) polegało na ustaleniu, czy metodologiczna jakość fokusa (rozumiana jak wyżej) przekłada się na jego jakość praktyczną (zysk informacyjny).

Pomiaru relatywnej siły aktywacyjnej analizowanych efektów grupowych dokonano na podstawie wskaźników obserwacyjnych i socjometrycznych, natomiast relatywne bogactwo informacyjne oceniano korzystając z procedur kodowania materiału jakościowego. Warto też wyjaśnić, dlaczego w niniejszym artykule posługujemy się pojęciem „relatywnej siły aktywacji efektów grupowych” oraz „relatywnego bogactwa informacyjnego”. Otóż nie jesteśmy w stanie wyznaczyć absolutnych progów określających zakres liczbowy, np. wysokiej siły aktywacji efektu śnieżnej kuli, czy niskiego stopnia unikalności zebranych informacji, gdyż niewielka liczba zrealizowanych w ramach projektu badawczego sesji FGI ( $n=5$ ) nie uprawnia do sporządzenia normalizacji dla globalnej metodologicznej i praktycznej jakości fokusa, jak i również poszczególnych wymiarów tych konstruktywów. Wypracowana procedura konkretyzacji efektów grupowych z uwagi na brak normalizacji nie upoważnia do wyrażenia stwierdzenia, że w danej grupie A określony efekt zaaktywizował się na niskim, przeciętnym lub wysokim poziomie. Uprawnia jednak do sformułowania spostrzeżenia, że stopień nasilenia danego efektu w określonej grupie jest wyższy, niższy albo porównywalny na tle innej grupy lub grup FGI.

Wymóg pogodzenia ilościowych wskaźników konkretyzujących siłę aktywacji efektów grupowych i jakościowych wskaźników konkretyzujących praktyczną jakość badania FGI osiągnięto na drodze zastosowania eksperymentu naturalnego polegającego na realizacji pięciu sesji FGI prowadzonych zgodnie z poczynionymi przez nas założeniami tradycyjnej i eksperymentalnej konwencji FGI (wątek ten zostanie dokładniej omówiony w dalszej części opracowania). Zachowując istotę badania FGI (tj. pozyskanie materiału na drodze generowania informacji przez wzajemnie oddziałujących na siebie dyskutantów) zachowano jego jakościowy status. Natomiast nałożenie na scenariusz merytoryczny określonych warunków odnoszących się do sposobu kierowania dyskusją przez moderatora (in. warunków eksperymentalnej konwencji FGI) zapewniło aranżację sytuacji badawczej umożliwiającej dokonanie zobiektywizowanych porównań międzygrupowych



w zakresie nasilenia efektów grupowych i wewnątrzgrupowych w zakresie częstości wybranych zachowań dyskutantów, np. częstości kierowania pytań do rozmówców.

W ramach naszego projektu badawczego zrealizowano 5 sesji fokusowych z udziałem młodzieży o porównywalnym kilkumiesięcznym stażu funkcjonowania w nowej dla siebie przestrzeni instytucjonalnej. Wybór tematu będącego przedmiotem dyskusji (obraz polskiego systemu edukacji w oczach młodzieży) nie był przypadkowy. Każdy z badanych dysponował osobistymi doświadczeniami nabytymi w procesie przystosowywania się do wymogów nowej roli, stąd każdy z nich występował w charakterze znawcy tematu (eksperta w sprawie) i w konsekwencji dysponował wiedzą umożliwiającą aktywne włączenie się do dyskusji. Najogólniej rzecz ujmując, dobór uczestników opierał się na kryterium merytorycznym (wspólnota doświadczeń edukacyjnych), demograficznym (zbliżony wiek osób badanych) i kryterium metodologicznym (wyniki uzyskane przez badanych w wymiarze konformizmu-nonkonformizmu). Dla uzyskania materiału umożliwiającego realizację zaplanowanych celów metodologicznych grupy były wewnętrznie homogeniczne i zewnętrznie zróżnicowane pod względem wymiaru osobowości konformizm – nonkonformizm. Wszystkie sesje były prowadzone przez tego samego moderatora i w tej samej przestrzeni (Pracownia Badań Fokusowych Instytutu Socjologii UŁ). Moderator posługiwał się scenariuszem, w którym uwzględniono płaszczyznę merytoryczną i metodologiczną. Dokładniejszy opis zastosowanych konwencji tradycyjnej i eksperymentalnej zamieszczamy w dalszej części artykułu.

Finalnym rezultatem opisywanego projektu badawczego było uporządkowanie pięciu zrealizowanych sesji na wymiarze globalnej jakości metodologicznej i na wymiarze globalnej praktycznej jakości FGI. Celem dokonania pomiaru globalnej jakości metodologicznej, każdą sesję oceniano na 6 wymiarach analitycznych określających relatywną siłę aktywacji efektu bezpieczeństwa, spontaniczności, stymulacji, śnieżnej kuli, konformizmu normatywnego oraz syndromu myślenia grupowego, poprzez przypisanie jej w obrębie każdego wymiaru rangi z zakresu 1–5. Rangę 1 przypisywano wtedy, gdy dana grupa FGI ulokowała się na tle czterech pozostałych na ostatnim miejscu pod względem stopnia nasilenia danego pozytywnego efektu grupowego (efektu bezpieczeństwa, spontaniczności, śnieżnej kuli, stymulacji) i wtedy, gdy dana grupa usytuowała się na pierwszym miejscu pod względem siły aktywacji negatywnego efektu grupowego (konformizmu, czy syndromu grupowego myślenia). W rezultacie globalna jakość metodologiczna fokusa jest różnicą między sumą rang przypisaną konkretnej grupie na czterech wymiarach nasilenia pozytywnych efektów (zakres

4–20 punktów, gdzie 20 oznacza najwyższą globalną siłę aktywacji pozytywnych efektów) a sumą rang przypisanych tej grupie na dwóch wymiarach nasilenia negatywnych efektów (zakres 2–8 punktów, gdzie 8 punktów oznacza najwyższą globalną siłą aktywacyjną efektów negatywnych). Zatem w przypadku 5 sesji maksimum dla metodologicznej jakości fokusa wynosi 18 punktów (działanie: 20–2), zaś minimum jest równe 4 punkty (działanie: 4–8). Jeśli przedmiotem analizy jest porównanie 5 sesji, to metodologiczna jakość fokusa przyjmuje wartości z zakresu –4 do 18 i im bliższa jest 18, tym jakość ta jest lepsza. Jeśli naszym celem jest porównanie 5 sesji pod kątem jakości praktycznej, to z kolei można zauważyć, iż jakość ta jest obliczana jako suma rang z zakresu 1–5 na trzech wymiarach analitycznych: różnorodności, unikalności i nasycenia pozyskanych informacji. Stąd, jakość merytoryczna może przyjmować wartości z zakresu 3–15 i im bliższa jest 15, tym jakość ta jest lepsza.

Na potrzeby badań przyjęliśmy w oparciu o literaturę, że do pozytywnych efektów grupowych zalicza się: efekt stymulacji, śnieżnej kuli, bezpieczeństwa oraz spontaniczności [Malinowski, 2007], a do negatywnych z kolei konformizm i syndrom grupowego myślenia [Oyster, 2002 za: Steiner 1972]. Warto w tym miejscu podkreślić, nieco wybiegając w przód, że, jak wykazały wyniki naszych badań, proponowany w literaturze podział na pozytywne oraz negatywne efekty grupowe w przypadku badań fokusowych wiąże się z pewnymi ograniczeniami. Dla przykładu – nadmiarowo silna aktywacja pozytywnych efektów emocjonalnych – efektu bezpieczeństwa i spontaniczności może stymulować dyskutantów do nawiązywania ze sobą luźnych, towarzyskich konwersacji i tym samym przyczynić się do tego, że pole dyskusji jest wypełniane wątkami dygresyjnymi, a nie wypowiedziami zgodnymi z intencją moderatora i pozwalającymi zaspokoić zapotrzebowanie informacyjne badacza. Innymi słowy nadmiarowa siła aktywacji efektów emocjonalnych może odciągać badanych od właściwego zadania, tj. generowania wypowiedzi, a także pogłębiania, uzupełnienia, komentowania lub dokonywania klaryfikacji poglądów podzielanych przez rozmówców.

#### **DYNAMIKA GRUPOWA – EFEKTY GRUPOWE A METODOLOGICZNA JAKOŚĆ FOKUSA**

Podstawową właściwością badania fokusowego jest fakt, że informacje generowane w trakcie sesji są wynikiem wspólnej pracy dyskutantów, stanowiąc skutek dynamiki grupowej, czyli całokształtu uświadomionych lub pozostających poza świadomością badanych interakcji zachodzących między nimi [Nikodem-

ska-Wołownik, 1999]. Na potrzeby własnych badań doprecyzowaliśmy powyższą definicję przyjmując, że interakcje te są modyfikowane przez właściwości strukturalne grupy. Przyjeliśmy, że interakcjami rozgrywającymi się w obrębie grupy fokusowej jest np. kierowanie pytań do innych dyskutantów (interakcja uświadomiona), werbalne potakiwania (interakcja uświadomiona lub zachowanie odruchowe), zaś do kategorii właściwości grupy można przykładowo zaliczyć stopień nasycenia atmosfery relacji wewnątrzgrupowych odczuciami sympatii, mierzony za pomocą własnego wskaźnika socjometrycznego. Z naszych badań wynika bowiem między innymi, że wyłonienie się w grupie dyrektywnej jednostki, kontestującej opinie innych i absolutyzującej słuszność własnych racji może doprowadzić do zawiązania się koalicji. Możliwe, że dyrektywny styl prezentacji poglądów może być interpretowany przez rozmówców jako zagrożenie dla spójności grupy lub zapędy do praktykowania autorytarnego stylu kierowania dyskusją i negacji słuszności konwencjonalnych poglądów. Stąd wyłonienie się w strukturze grupowej roli dyrektywnego opozycjonisty może stymulować pozostałych jej członków do jednoczenia się w koalicję, a rezultatem ścierania się przeciwstawnych opinii może być, np. utrzymywanie wysokiego pobudzenia w grupie i w efekcie uzyskanie większej dawki informacji. Z kolei nadmiarowe nasycenie atmosfery kontaktów wewnątrzgrupowych odczuciami sympatii może być siłą wstrzymującą badanych przed werbalizacją odmiennych przekonań, gdyż odmienne opinie mogą być przyczynkiem do konfliktu, spadku spójności grupowej i zaostrzenia obaw i niepokojów wiążących się z uczestnictwem w danej grupie.

Prawidłowa dynamika grupowa powinna odznaczać się poniższymi właściwościami:

- odpowiednim poziomem aktywności poszczególnych dyskutantów – nie za dużym, by uniknąć monopolizacji dyskusji przez pojedyncze osoby i nie za małym, by zapobiegać utracie informacji wskutek ich wycofywania się,
- dominacją relacji badany – grupa nad relacją badany – moderator, by nie stracić charakteru zogniskowanego wywiadu grupowego na rzecz wywiadów indywidualnych,
- inicjatywą uczestników w podejmowaniu polemiki z pozostałymi osobami lub inicjatywą polegającą na wspieraniu opinii interlokutorów dodatkowymi argumentami,
- swobodną ekspresją odczuwanych emocji,
- życzliwą postawą uczestników spotkania wobec pozostałych badanych, zachęcającą ich do zabierania głosu w dyskusji.

W literaturze podkreśla się, że bezpośrednim następstwem dynamiki grupowej są efekty grupowe [Malinowski, 2007]. Można zatem wnioskować, że efekty grupowe są następstwem świadomych lub nieświadomych interakcji rozgrywających się między dyskutantami, a zatem próba konkretyzacji siły aktywacji negatywnych i pozytywnych efektów grupowych na podstawie wymiarów funkcjonowania dyskutanta, np. werbalizacji reakcji potakujących (in. wymiarów opisujących konkretne interakcje zachodzące między dyskutantami) wydaje się zasadna z teoretycznego punktu widzenia. Co więcej integrując 2 poniższe przesłanki: 1) metodologiczna jakość fokusa jest ujmowana jako stopień nasilenia efektów grupowych, 2) efekty grupowe są bezpośrednim następstwem dynamiki grupowej, można stwierdzić, że metodologiczna jakość FGI jest dokonywaną przez badacza oceną poprawności dynamiki grupowej. Dynamika ta jest tym bardziej prawidłowa, im rejestruje się większą siłą aktywacji pozytywnych efektów grupowych i mniejsze nasilenie efektów negatywnych z zastrzeżeniem, że nadmiarowa siła aktywacji efektów emocjonalnych może skutkować redukcją nasilenia efektu stymulacji oraz śnieżnej kuli. Ocena poprawności dynamiki grupowej powinna zatem uwzględniać przejawy tej dynamiki, czyli pozytywne i negatywne efekty grupowe, a ściślej ich siłę aktywacyjną. Z kolei ocena stopnia nasilenia wybranych efektów powinna bazować na częstości obserwowalnych zachowań dyskutantów, np. częstości artykulacji odmiennych przekonań. Mamy nadzieję, że powyższe rozważania w wystarczającym stopniu ukazałyśmy dysfunkcje zaznaczające się między pojęciami metodologicznej jakości FGI, dynamiki grupowej i efektów grupowych. Zatem dynamika grupowa i jej następstwa – efekty grupowe – są „dziełem” grupy, a metodologiczna jakość FGI (podobnie jak jakość praktyczna) jest konstruktem tworzonym przez badacza, który opisuje poprawność dynamiki grupowej oraz stopień nasilenia efektów grupowych.

### **PROCEDURA OCENY METODOLOGICZNEJ JAKOŚCI FOKUSA**

Poniżej prezentujemy proponowany przez nas sposób oceny metodologicznej jakości fokusa. Analizę rozpoczniemy od przedstawienia propozycji diagnozy nasilenia pozytywnych efektów grupowych, przechodząc dalej do efektów negatywnych.

Przyjętą przez nas jednostką analizy jest każda wypowiedź dyskutanta wygenerowana w toku danej sesji badawczej. Przy tym za wypowiedź uznaliśmy każdą reakcję werbalną, poczynając od pojedynczego słowa, a kończąc na rozbudowanej narracji oraz każdy komunikat paralingwistyczny (np. mhm, aha) interpretowany

kontekstualnie jako wyraz potwierdzenia lub zaprzeczenia. Ponadto w sytuacji, gdy wypowiedź dyskutanta zostaje przerwana przez niego lub innego rozmówcę na okres nie dłuższy niż trzy sekundy i jest kontynuowana, traktujemy to ją jako jedną jednostkę.

Celem operacjonalizacji 4 pozytywnych efektów grupowych (efektu bezpieczeństwa, spontaniczności, stymulacji, śnieżnej kuli) zastosowaliśmy procedurę sędziów kompetentnych z udziałem wykładowców z Katedry Metod i Technik Badawczych Instytutu Socjologii UŁ. Zadanie sędziów polegało na oszacowaniu na pięcio-stopniowej skali stopnia, w jakim dany wskaźnik charakteryzujący wybrany aspekt dynamiki badania fokusowego pozwala na pomiar każdego z wyżej wymienionych efektów grupowych. Założyliśmy, że ten sam wskaźnik może cechować więcej niż jeden efekt, przy czym stopień, w jakim wskaźnik naznacza różne efekty może być zróżnicowany. Sędziowie szacowali adekwatność powiązania wskaźnika z efektem na 5-stopniowej skali: 1 – zdecydowanie nie opisuje, 2 – raczej nie opisuje, 3 – trudno powiedzieć, 4 – raczej opisuje albo 5 – zdecydowanie opisuje. Przyjęto, że wskaźnik pozwala na pomiar nasilenia danego efektu grupowego, jeżeli dla konkretnej kombinacji wskaźnika i efektu co najmniej 4 spośród 7 sędziów wybrało odpowiedź „raczej opisuje” lub „zdecydowanie opisuje”. Sędziowie oceniali 9 wskaźników: całkowity czas upływający między następującymi po sobie wypowiedziami kolejnych badanych, całkowity czas upływający od chwili zadania pytania przez moderatora a pojawieniem się pierwszej reakcji werbalnej ze strony grupy, liczbę: reakcji śmiechu, pytań skierowanych do innych dyskutantów, wypowiedzi odmiennych od opinii poprzedników, badanych reagujących werbalnie na pytania moderatora, wypowiedzi udzielonych przez całą grupę na pytania moderatora, wtrąceń i bezpośrednich nawiązań do wypowiedzi rozmówców. Sędziowie oceniali związek wymienionych wskaźników z czterema zdefiniowanymi na potrzeby tej procedury efektami grupowymi:

– efekt stymulacji – „ujawnia się w interakcjach uczestników spotkania, wypowiedź każdego członka dociera do innych i pobudza ich do działania” [Malinowski, 2007],

– efekt spontaniczności – ujawnia się w udzielaniu szybkich, odruchowych reakcji werbalnych oraz uniezależnieniu formułowanych poglądów (zarówno ich treści i formy) od realnych lub wyobrażonych oczekiwań innych członków grupy,

– efekt śnieżnej kuli – „ujawnia się w narastaniu informacji wokół jakiegoś wątku na skutek dołączania do niego komentarzy innych osób” [Malinowski, 2007],

– efekt bezpieczeństwa – manifestuje się w uogólnionym dla grupy poczuciu funkcjonowania w przestrzeni odczuwanej jako bezpieczna.

Z uwagi na to, że różne wskaźniki mogą w zróżnicowanym stopniu mierzyć ten sam efekt, obliczyliśmy dla wszystkich istotnych kombinacji: wskaźnik – efekt średnią arytmetyczną. Konstrukcję wzoru wyrażającego orientacyjne nasilenie efektu oparto na sumie iloczynów: średniej dla danej kombinacji wskaźnik – efekt oraz rangi definiującej miejsce, na którym konkretna grupa FGI ulokowała się na tle pozostałych grup pod względem wielkości określonego wskaźnika. Z uwagi na to, że zrealizowano 5 sesji, ranga ta przyjmuje wartości z zakresu 1–5, gdzie 5 oznacza wielkość wskaźnika najkorzystniejszą dla dynamiki badania.

Oszacowania dokonane przez sędziów kompetentnych pozwoliły wypracować podane niżej wzory określające orientacyjne nasilenie danego efektu:

$$ST = \sum (4,4 \bullet W_1) + (3,4 \bullet W_2) + (3,4 \bullet W_3) + (3 \bullet W_4) + (4,4 \bullet W_5) + (3,6 \bullet W_6) + (3,3 \bullet W_7) \quad (1)$$

$$(3,1 \bullet W_1) + (3,1 \bullet W_3) + (3,9 \bullet W_4) + (3,7 \bullet W_5) + (4,3 \bullet W_6) + (3,7 \bullet W_8) \quad (2)$$

$$\acute{S}K = \sum (3,9 \bullet W_1) + (3,1 \bullet W_2) + (3,4 \bullet W_5) + (3,4 \bullet W_6) \quad (3)$$

$$\sum (3,1 \bullet W_6) + (3,3 \bullet W_2) + (4,1 \bullet W_8) + (3,3 \bullet W_9) \quad (4)$$

gdzie:

ST – efekt stymulacji,

SP – efekt spontaniczności,

ŚK – efekt śnieżnej kuli,

B – efekt bezpieczeństwa,

R – miejsce, na jakim dana grupa fokusowa ulokowała się na danym wymiarze.

Indeksy dolne wprowadzono celem oznaczenia poszczególnych wskaźników:

1) odsetek komentarzy bezpośrednio nawiązujących do opinii rozmówców, czyli ich liczba podzielona przez  $N_c$  – całkowitą liczbę wypowiedzi wygenerowanych przez grupę jako całość na przestrzeni całego badania. Przy wyliczaniu nawiązań uwzględniono te, które zaczynały się od powtórzenia ostatnich słów poprzednika, ale jednocześnie zawierały nowe treści lub bezpośrednio korespon-

dowały z wypowiedziami rozmówców, np. *jak powiedział Michał, jak wcześniej zauważono, ja też mam podobne doświadczenia, mnie to również kojarzy się z... podobnie jak Kasia...*,

2) liczba wypowiedzi wygenerowanych przez grupę jako całość na 3 pytania moderatora z konwencji eksperymentalnej,

3) uśredniony dla konwencji eksperymentalnej odsetek badanych zabierających głos – wartość wskaźnika wyliczano w podany sposób: dla każdego z 3 bloków tematycznych (tj. 3 części eksperymentalnej konwencji fokusa) oddzielnie określano liczbę rozmówców przynajmniej raz zabierających głos, sumowano te wartości, a uzyskaną sumę dzielono przez maksymalną teoretyczną liczbę aktywnych rozmówców – iloczyn liczby części eksperymentalnej konwencji fokusa ( $n=3$ ) oraz liczebności grupy ( $N$ ), a wynik ilorazu mnożono razy 100%,

4) odsetek reakcji śmiechu – liczba tych reakcji podzielona przez  $N_c$ ,

5) odsetek pytań skierowanych do rozmówców – liczba tych pytań podzielona przez  $N_c$ ; przy liczeniu pytań, uwzględniono wyłącznie te, które spowodowały rozmówców do udzielenia na nie odpowiedzi,

6) odsetek wtrąceń – liczba wejść w słowo podzielona przez  $N_c$ ; uwzględniono wszystkie wtrącenia (tzn. wejścia w słowa) niezależnie, czy wypowiedź dyskutanta A została przerwana reakcją badanego B chwilowo, czy nieodwracalnie; w przypadku wtrącenia potakującego (np. *dokładnie*) lub opozycyjnego (np. *nie, tak to nie działa*) kwalifikowano wypowiedzi do obu kategorii: 1) do wtrąceń i 2) do wypowiedzi potakujących albo opozycyjnych, zaś pozostałe wtrącenia kwalifikowano wyłącznie do jednej kategorii,

7) odsetek wypowiedzi opozycyjnych – liczba odmiennych przekonań podzielona przez  $N_c$ ; uznaliśmy, że wypowiedź opozycyjna nie może się ograniczać do komunikatu, typu *nie* lub *nie, bo nie* (pozbawionego uzasadnienia); nie musi koniecznie zawierać formalnego zwrotu (np. *nie zgadzam się z tym, mam inne zdanie na ten temat, niekoniecznie tak jest*), ale jej treść wskazuje, że ukryta w niej intencja jest odmienna od przekonania wyłaniającego się z opinii rozmówcy/rozmówców;

8) łączny czas upływający między następującymi po sobie opiniami kolejnych dyskutantów podzielony przez  $T_k$ , tj. czas trwania konwencji eksperymentalnej badania FGI wyrażony w sekundach

9) całkowity czas upływający od momentu zadania pytania przez moderatora, a pojawieniem się pierwszej reakcji werbalnej ze strony dowolnego dyskutanta, podzielony przez  $T_k$ .

W poniższej tabeli przedstawiono wypracowaną na podstawie procedury sędziów kompetentnych konkretyzację efektów grupowych oraz średnią orienta-

cyjnie określającą, w jakim stopniu dany wskaźnik pozwala na pomiar nasilenia danego efektu.

TABELA 1. Pomiar pozytywnych efektów grupowych

| Lp. | WSKAŹNIK   | EFEKT GRUPOWY  |
|-----|--|--|
| 1   | Liczba bezpośrednich nawiązań  | Efekt stymulacji (4,4)<br>Efekt śnieżnej kuli (3,9)<br>Efekt spontaniczności (3,1)                               |
| 2   | Łączna liczba wypowiedzi udzielonych przez grupę na pytania moderatora   | Efekt stymulacji (3,4)<br>Efekt śnieżnej kuli (3,1)  |
| 3   | Liczba badanych udzielających odpowiedzi na pytania moderatora   | Efekt stymulacji (3,4)   |
| 4   | Liczba reakcji śmiechu   | Efekt spontaniczności (3,9)<br>Efekt stymulacji (3,0)  |
| 5   | Liczba pytań skierowanych do innych dyskutantów  | Efekt stymulacji (4,4)<br>Efekt spontaniczności (4,3)<br>Efekt śnieżnej kuli (3,4)                               |
| 6   | Liczba wtrąceń   | Efekt spontaniczności (4,3)<br>Efekt stymulacji (3,6)<br>Efekt śnieżnej kuli (3,4)<br>Efekt bezpieczeństwa (3,1) |
| 7   | Liczba wypowiedzi opozycyjnych   | Efekt stymulacji (3,3)<br>Efekt bezpieczeństwa (3,3)   |
| 8   | Czas upływający między następującymi po sobie wypowiedziami kolejnych dyskutantów  | Efekt bezpieczeństwa (4,1)<br>Efekt spontaniczności (3,7)  |
| 9   | Czas upływający od momentu zadania pytania przez moderatora, a pojawieniem się pierwszej reakcji werbalnej ze strony grupy | Efekt bezpieczeństwa (3,3)   |

Źródło: opracowanie własne

Następnie uzyskane wyniki zostały skonsultowane z praktykami badań FGI. Praktycy uznali, że wybory sędziów, czyli ustalone przez nich relacje między wskaźnikiem a efektem są trafne. W ten sposób ugruntowano trafność propozycji konkretyzacji pozytywnych efektów grupowych.

Celem oszacowania relatywnego nasilenia efektu stymulacji w grupach FGI, zastosowano zestaw 7 wskaźników: odsetek wypowiedzi bezpośrednio nawiązujących do opinii innych rozmówców ( $W_1$ ), liczba wypowiedzi wygenerowanych przez grupę jako całość podczas eksperymentalnej konwencji badania ( $W_2$ ), uśredniony dla całej konwencji eksperymentalnej odsetek badanych zabierających głos ( $W_3$ ), odsetek reakcji śmiechu ( $W_4$ ), pytań skierowanych do innych dyskutantów ( $W_5$ ), wtrąceń ( $W_6$ ) i odmiennych przekonań ( $W_7$ ). Decyzją



sędziów kompetentnych, za pomocą tych wskaźników można ocenić nasilenie efektu stymulacji przejawiającego się w tym, że wypowiedź danego dyskutanta dociera do innych osób, wyzwalając w nich motywację do działania, a ściślej do werbalizacji produktów własnej aktywności umysłowej, np. opinii, czy skojarzeń [Lisek-Michalska, 2007].

Z kolei relatywne nasilenie efektu śnieżnej kuli w grupach fokusowych, którego istotą jest generowanie informacji w oparciu o posiadane dane [Malinowski, 2007] sędziowie kompetentni oszacowali na podstawie zestawu 4 wskaźników takich jak: odsetek bezpośrednich nawiązań ( $W_1$ ), łączna liczba wypowiedzi wygenerowanych przez grupę podczas eksperymentalnej konwencji badania ( $W_2$ ), odsetek pytań skierowanych do rozmówców ( $W_3$ ) i wejść w słowo ( $W_6$ ). Uzyskane rozstrzygnięcie wskazujące, że wszystkie 4 wskaźniki efektu śnieżnej kuli zostały również przyporządkowane do efektu stymulacji może sugerować, że istnieje ścisła relacja między tymi efektami. Siła aktywacji efektu śnieżnej kuli może świadczyć o poziomie kooperacji w grupie (ibidem). Warunkiem inicjacji współpracy przejawiającej się w dobudowywaniu do generowanych wypowiedzi nowych komentarzy jest angażowanie zasobów uwagi dowolnej, a to z kolei wymaga wzbudzenia i podtrzymywania motywacji zadaniowej. I tak, dostatecznie silną aktywację efektu stymulacji można traktować jako wyjściowy warunek wystąpienia efektu śnieżnej kuli, a oba te efekty można w naszej opinii odnieść do zadaniowego wymiaru funkcjonowania dyskutantów.

Z kolei celem dokonania oceny relatywnego nasilenia efektu bezpieczeństwa, czyli uogólnionego dla grupy poczucia funkcjonowania w przestrzeni sytuacyjnej i interakcyjnej wzbudzającej komfort psychiczny, sędziowie zaproponowali zestaw 4 wskaźników: odsetek wtrąceń ( $W_6$ ) i odmiennych opinii ( $W_7$ ), łączny czas upływający między następującymi po sobie wypowiedziami kolejnych rozmówców ( $W_8$ ) i całkowity czas upływający od momentu zadania pytania przez moderatora a pojawieniem się pierwszej reakcji werbalnej dowolnego rozmówcy ( $W_9$ ).

Natomiast w celu oceny relatywnej siły aktywacyjnej efektu spontaniczności, którego istotą ujawnia się w udzielaniu szybkich, odruchowych reakcji werbalnych i uniezależnieniu treści lub formy artykułowanych opinii od rzeczywistej lub wyobrażonej presji oczekiwań innych członków grupy, sędziowie zaproponowali zestaw 6 podanych wskaźników: odsetek bezpośrednich nawiązań do wypowiedzi rozmówców ( $W_1$ ), uśredniony dla eksperymentalnej konwencji odsetek dyskutantów zabierających głos ( $W_3$ ), odsetek reakcji śmiechu ( $W_4$ ), pytań skierowanych do innych badanych ( $W_5$ ), wtrąceń ( $W_6$ ) oraz łączny czas upływający między następującymi po sobie wypowiedziami kolejnych dyskutantów ( $W_8$ ).

Brak kompletnych założeń teoretycznych odnoszących się do istoty fokusów, utrudnia wyszczególnienie atrybutów „dobrego badania FGI” oraz powoduje, że każda taka propozycja jest obciążona dużą dozą subiektywizmu. W naszym odczuciu lepsza jakość metodologiczna fokusa wiąże się z wyższym nasileniem efektu stymulacji, śnieżnej kuli, spontaniczności, jak i bezpieczeństwa oraz niższym nasileniem grupowego myślenia i konformizmu normatywnego. Uwzględniając wskaźniki pozytywnych efektów grupowych można sprecyzować, że lepszy jest fokus, w którym występuje większa liczba: bezpośrednich nawiązań, wypowiedzi wygenerowanych przez grupę na pytania moderatora, dyskutantów werbalizujących swoje przekonania, reakcji śmiechu, pytań do innych rozmówców, wtrąceń, odmiennych poglądów, a także krótszy czas upływający od momentu zadania pytania przez moderatora, a pierwszą reakcją werbalną grupy i krótszy czas między następującymi po sobie wypowiedziami kolejnych rozmówców. Należy jednakże podkreślić, że nadmiarowa liczba wtrąceń, reakcji śmiechu i wypowiedzi opozycyjnych może uczynić z badania FGI arenę nieustannych potyczek słownych; ścierania się chaotycznych, niedokończonych wypowiedzi lub luźne spotkanie towarzyskie przepełnione wypowiedziami dygresyjnymi. Ważne staje się to, że nadmiar określonych reakcji korzystnych dla przebiegu badania FGI może optymalizować tylko stopień nasilenia niektórych efektów (np. efektów emocjonalnych) kosztem spadku siły aktywacji innych pozytywnych efektów lub też wzrostu nasilenia efektów negatywnych. Odwołując się do konkretnego przykładu można zauważyć, że zbyt duża liczba reakcji śmiechu w grupie może prowadzić do nadmiarowej aktywacji efektów emocjonalnych, a to z kolei może wiązać się z nadaktywacją syndromu grupowego myślenia i słabszym nasileniem efektu stymulacji i śnieżnej kuli.

Celem dokonania pomiaru nasilenia poszczególnych efektów grupowych, na etapie projektowania scenariusza przyjęto kilka istotnych założeń metodologicznych. Na scenariusz merytoryczny nałożono schemat tradycyjnej i eksperymentalnej konwencji fokusowej. Dyskusję dotyczącą wyobrażeń badanych, postrzeganych przez nich korzyści, strat i obrazu siebie w nowej roli społecznej prowadzono zgodnie z zaleceniami tradycyjnej konwencji FGI. Z kolei założenia konwencji eksperymentalnej FGI zastosowano w technice chińskiego portretu (dyskutanci porównywali szkołę do zwierzęcia i marki samochodu) i bloku 3 pytań zasadniczych (dyskutanci dzielili się poglądami, nt. znajomości nawiązanych w nowej szkole, swoich nauczycieli, dodatkowych możliwości oferowanych przez szkołę).

Owa tradycyjna konwencja bazuje na założeniu, że moderator dysponuje pełną możliwością dopasowywania stylu kierowania dyskusją do poziomu aktywności

werbalnej grupy i klimatu panującego w niej. Przekładając to ogólne założenie na język szczegółowych zaleceń, ustalono, iż w konwencji tradycyjnej moderator może operować klaryfikacjami, pytaniami pogłębiającymi lub podsumowaniami, tak często, jak uzna to za potrzebne. Kluczowym zadaniem moderatora, specyficznym dla konwencji tradycyjnej było dopytywanie się wszystkich dyskutantów nie zabierających głosu z własnej inicjatywy. Celem tego zabiegu było podjęcie próby wdrukowania badanym reguły dzielenia się własnymi spostrzeżeniami, uwrażliwienia liderów na obecność milczków oraz przełamywania oporów milczków przed publiczną artykulacją własnych poglądów.

Jeśli zaś chodzi o konwencję eksperymentalną, to przyjęliśmy, że pierwszym jej warunkiem jest to, że moderator nie zaprasza do udziału w dyskusji poprzez wywołanie do odpowiedzi poszczególnych dyskutantów lecz zwraca się do grupy jako całości. Aktywność werbalna badanego jest regulowana przez niego samego i innych rozmówców. Drugim warunkiem polega na tym, że moderator jest upoważniony do formułowania pytań tylko na poziomie grupy (pytania typu, co myślicie na temat...), a nie jednostki. Trzeci sygnalizuje, że jeśli po 5 sekundach po zakończeniu wypowiedzi, żaden z badanych nie ujawni jakiegokolwiek reakcji werbalnej, to moderator jest zobowiązany do zadania pytania: „czy jest coś jeszcze, czym chcielibyście się podzielić z grupą?”. I jeśli na przestrzeni najbliższych pięciu sekund nie zarejestruje się żadnej reakcji werbalnej ze strony dyskutantów, to moderator formułuje następne pytanie rozpoczynające nowy wątek tematyczny. Jeżeli reakcja werbalna zostanie ujawniona przez jakiegokolwiek badanego, to dyskusja jest kontynuowana, aż do tego momentu, gdy ponownie czas ciszy przekroczy 5 sekund – wtedy procedura jest powtarzana. Czwarty warunek tej konwencji jest uzupełnieniem poprzedniego i głosi, że jeżeli w ciągu 10 sekund od momentu zakończenia artykulacji pytania, moderator nie otrzyma żadnej odpowiedzi, to parafrazuje treść pytania. Jeśli zaś po upływie kolejnych 5 sekund, nie zarejestruje się żadnej reakcji werbalnej ze strony dyskutantów, to przechodzi do następnego pytania. Piąte założenie wskazuje, że moderator jest zobligowany do ścisłego przestrzegania zaplanowanego łącznego czasu trwania dyskusji prowadzonej w obrębie konkretnego wątku. Ustalono, że maksymalny czas trwania dyskusji utrzymanej w konwencji eksperymentalnej wynosi kolejno ok. 13 minut dla pierwszego wątku tematycznego, zaś dla dwóch następnych wątków po ok. 12 minut. Progi te ustalono na podstawie obserwacji przebiegu pilotażowej sesji. Szósty warunek stanowiący uszczegółwienie poprzedniego wskazuje, że dyskusja w obrębie danego wątku może zostać sfinalizowana przez moderatora po tym, jak ostatni dyskutant (zabierający głos jeszcze przed upływem maksymalnego progu czasu trwania dyskusji) zakończy swoją wypowiedź.

Siódmy warunek sygnalizuje zaś, że moderator w konwencji eksperymentalnej nie może stosować klasyfikacji i pytań pogłębiających, gdyż poprzez operowanie takimi zabiegami nie odnosiłby się do grupy, jako całości, lecz do jej konkretnych członków. Przyjęto, że moderator podczas trwania konwencji tradycyjnej i eksperymentalnej powinien okazywać dyskutantom zainteresowanie poprzez utrzymywanie kontaktu wzrokowego i sygnały paralingwistyczne.

Potrzebę zastosowania zaleceń konwencji eksperymentalnej można uzasadnić tym, że poprzez stosowanie zabiegów typowych dla konwencji tradycyjnej, tj. klasyfikacji, pytań pogłębiających lub parafraz – moderator może albo wyrównywać poziom aktywności werbalnej poszczególnych dyskutantów albo nieświadomie faworyzować wybranych rozmówców, udzielając więcej oznak uwagi i zainteresowania dla ich wypowiedzi. Tą przyczyną owej nieświadomej faworyzacji może być, np. atrakcyjność fizyczna dyskutantów, postrzegana przez moderatora zbieżność między osobistymi opiniami, a poglądami badanych, szczególnie cennie przez niego cechy osobowości, temperamentu lub style narracji. Dlatego w konwencji eksperymentalnej zakres możliwych interwencji moderatora został zawężony do zadawania pytań uwzględnionych w scenariuszu, monitorowania czasu trwania dyskusji w obrębie wątku oraz utrzymywania kontaktu wzrokowego z rozmówcami. Powyższe założenia korespondują z idealnym modelem fokusa jako badania napędzanego siłami motywacji samych dyskutantów oraz badania, w którym występuje wyraźne odgraniczenie roli moderatora od roli rozmówcy. Naszym zdaniem wartości 4 spośród 9 zaproponowanych przez nas wskaźników opisujących dynamikę badania FGI są szczególnie czułe na wpływ zróżnicowanej aktywności moderatora i dlatego wyliczono je tylko dla bloków tematycznych moderowanych zgodnie z warunkami konwencji eksperymentalnej. Do tej grupy wskaźników zaliczono:

- 1) całkowity czas upływający między następującymi po sobie wypowiedziami kolejnych badanych
- 2) łączny czas upływający między momentem zadania pytania przez moderatora a aktywacją pierwszej reakcji werbalnej ze strony grupy – moderatorzy różnią się w zakresie tolerancji co do czasu trwania ciszy
- 3) liczbę wypowiedzi wygenerowanych przez grupę na pytania moderatora – im większa aktywność moderatora w formułowaniu pytań, tym wyższe prawdopodobieństwo wygenerowania większej liczby wypowiedzi przez grupę jako całość
- 4) liczbę badanych zabierających głos w dyskusji – moderator może zachęcać milczków do włączenia się do dyskusji lub „akceptować” ich pasywność.

O ile celem etapu wstępnej aranżacji i tradycyjnej konwencji było wdrukowanie uczestnikom reguły wskazującej, że podstawowym elementem roli uczestnika badań FGI jest dzielenie się z innymi własnymi spostrzeżeniami, to z kolei założeniem konwencji eksperymentalnej było wyraźne odgraniczenie roli moderatora od roli rozmówcy i minimalizacja wpływu moderatora na przebieg badania. Zawężenie aktywności moderatora w konwencji eksperymentalnej i przekazanie informacji badanym o regułach działania grup FGI pozwoliło zobjektywizować w wymiarze międzygrupowym pomiar siły aktywacji wybranych pozytywnych efektów grupowych.

Z kolei wielkości pozostałych wskaźników są naszym zadaniem w większym stopniu zależne od właściwości badanych lub właściwości grupy. Do tej kategorii wskaźników zaliczono liczbę: bezpośrednich nawiązań, reakcji śmiechu, wtrąceń, wypowiedzi opozycyjnych oraz pytań skierowanych do rozmówców. Wartości tych wskaźników ustalano na podstawie materiału pozyskanego z całego przebiegu badania FGI. Z uwagi na większą odporność 5 powyższych wskaźników na wpływ moderatora i przyjętego przez niego stylu kierowania dyskusją, wartości tych wskaźników wyliczono też na podstawie materiału uzyskanego z tradycyjnej konwencji badania.

W przypadku negatywnych efektów grupowych proponujemy pomiar siły aktywacji dwóch negatywnych efektów grupowych: syndromu myślenia grupowego oraz konformizmu grupowego. Nie interesuje nas natomiast pomiar próżniactwa społecznego rozumianego jako zjawiska, w którym obecność innych generuje spadek produktywności działania jednostki i ujawniającego się szczególnie, gdy wykonywane zadanie jest dobrze wyuczone, postrzegane przez jednostkę jako mniej ważne lub gdy trudniej dokonać oceny indywidualnego wkładu pracy w realizację grupowego zadania [Domachowski, 2007]. Decyzję o pominięciu pomiaru nasilenia próżniactwa społecznego motywujemy tym, że lider aktywności werbalnej – osoba najczęściej zabierająca głos podczas dyskusji może operować różnorodnymi argumentami, zwiększając tym samym liczbę kategorii odpowiedzi wyodrębnionych podczas analizy materiału albo permanentnie powielać wciąż te same opinie. W tej drugiej sytuacji, próżniactwo może przekładać się na uzyskanie mniejszego bogactwa informacyjnego, gdyż lider przejawia skłonność do reprodukcji tych samych treściowo opinii, a pozostali dyskutanci relatywnie rzadko zabierają głos. Ponadto równomierny rozkład czasu przeznaczanego na wypowiedź każdego dyskutanta nie musi być równoznaczny z bogactwem informacyjnym otrzymanego materiału. Może zdarzyć się tak, że liczba kategorii wypowiedzi w grupie zdominowanej przez pojedynczych dyskutantów jest

większa od liczby kategorii odpowiedzi wygenerowanych w grupie, w której wszyscy dyskutanci byli porównywalnie zaangażowani w dyskusję.

Zrezygnowaliśmy także z pomiaru siły aktywacji efektu polaryzacji, rozumianej jako tendencja do podejmowania bardziej radykalnych decyzji i prezentacji bardziej ekstremalnych opinii w grupie aniżeli indywidualnie [Oyster, 2002]. Nasza decyzja wynika, m.in. z faktu, że w literaturze brakuje argumentów uzasadniających status efektu polaryzacji (in. przesunięcia ryzyka) jako negatywnego efektu grupowego. Ponadto wyszliśmy z założenia, że polaryzacja definiowana jest w oparciu o zmianę w zakresie radykalności opinii, a nie jej treści. W związku z tym zakładając, że zjawisko to uaktywniłoby się podczas trwania badania FGI, to prawdopodobnie nie miałyby wpływu na treściową różnorodność sformułowanych opinii, lecz bardziej na stopień ich wyrazistości.

Konformizm jako zjawisko bada się, korzystając z procedur eksperymentalnych (np. seria eksperymentów Ascha lub Scheriffa) ufundowanych na założeniu, że eksperymentator poprzez manipulację nasileniem zmiennych niezależnych (najczęściej sytuacyjnych, rzadziej osobowościowych) prowokuje jego aktywację [Bernacka, 2004; Tarkowska 1999]. Z jednej strony, procedura eksperymentalna (pod warunkiem, że uwzględnia randomizację) pozwala na ustalenie związków przyczynowo-skutkowych między zmiennymi, jednakże często nie spełnia wymogu realizmu sytuacyjnego. Potrzeba sformułowania zaleceń dotyczących optymalizacji jakości badania FGI i postulat zachowania realizmu sytuacji skłoniły nas do wnioskowania o sile konformizmu głównie na podstawie wskaźników obserwacyjnych. Przyjeliśmy, że w grupach FGI istota konformizmu manifestuje się w podporządkowaniu się do norm i oczekiwań grupy, a normą tą (a raczej metanormą) jest przemilczanie odmiennych przekonań. Ponadto na drodze estymacji modelu regresji liniowej dla częstości werbalizacji wypowiedzi opozycyjnych wykazano, że istotnymi predyktorami tej zmiennej są: nasilenie emocji adaptacyjnych odczuwanych przez dyskutanta podczas badania FGI, wymiar zachowania algorytmiczne – heurystyczne (wymiar preferencji intelektualnych) oraz nasilenie emocji kontrolujących aktywowanych na myśl o przystąpieniu do dyskusji. Powyższe zmienne objaśniające pozwalają wyjaśnić ok. 35% wariacji częstości artykulacji odmiennych opinii ( $R^2=0,35$ ). Precyzując, im wyższe nasilenie emocji adaptacyjnych (np. zapалу, ożywienia, radości), silniejsze preferencje heurystycznego stylu myślenia, a także niższe nasilenie emocji kontrolujących (np. lęku przed utratą akceptacji grupy, poczucia skrępowania), tym częstsza skłonność do artykulacji odmiennych przekonań w obecności członków własnej grupy fokusowej. Można zatem sądzić, że nasilenie emocji kontrolujących nie jest jedyną zmienną wpływającą na częstość werbalizacji odmiennych przekonań.

Badana przez nas tendencja do powstrzymywania się przed publicznym wygłoszeniem odmiennego przekonania jest regulowana zarówno przez dynamicznie zmieniające się w czasie zmienne afektywne, jak i bardziej stabilne preferencje intelektualne ujmowane na wymiarze zachowania algorytmiczne – heurystyczne.

Pierwszy wskaźnik cząstkowy niezbędny do ustalenia siły konformizmu – percepcja rozbieżności opinii (WRO) jest opracowywany z danych socjometrycznych dla 2 pomiarów, dla których odstęp czasowy wynosił w przybliżeniu ok. 60 minut. Za pomocą wskaźnika można orientacyjnie oszacować, jak silna (na tle innych grup) jest uogólniona tendencja do postrzegania przekonań rozmówców jako odmiennych od własnych.

Oblicza się go następująco:

$$WRO_x = \frac{\left(\frac{nWN_1}{nW_1} \cdot 100\%\right) + \left(\frac{nWN_2}{nW_2} \cdot 100\%\right)}{nP}$$

gdzie:

- x – nr grupy fokusowej,
- nWN – liczba odrzuceń udzielonych przez grupę na wymiarze III,
- nW – łączna liczba wyborów na wymiarze III,
- nP – liczba pomiarów (tutaj: 2),
- 1 – pretest,
- 2 – postest.

Im wyższa wartość % WRO, tym silniejsza uogólniona dla całej grupy tendencja do percepcji opinii innych dyskutantów jako odmiennych od własnych, osobistych przekonań. Kolejny wskaźnik cząstkowy (WO) informuje, jaki udział % w ogólnej liczbie wypowiedzi wygenerowanych podczas dyskusji reprezentują wypowiedzi opozycyjne. Wskaźnik ten jest ilorazem liczby wypowiedzi opozycyjnych do łącznej liczby wypowiedzi wygenerowanych przez grupę na przestrzeni dyskusji, pomnożonym razy 100%. Kolejny wskaźnik ( $W_{15}$ ) – powstrzymywanie się od publicznej ekspresji wypowiedzi opozycyjnych – jest ilorazem WRO do WO. Im wyższa tendencja do postrzegania przekonań rozmówców jako rozbieżnych z własnymi (wyższa wartość WRO) oraz im rzadziej w grupie artykułuje się wypowiedzi opozycyjne (niższa wartość WO), tym silniejsza tendencja do przemilczania odmiennych przekonań (niższa wartość  $W_{15}$ ). Do pomiaru siły aktywacyjnej konformizmu zastosowano także wskaźnik  $W_{14}$ , tzn. stosunek liczby wypowiedzi potakujących do opozycyjnych. Wskaźnik ten użyto również

do pomiaru nasilenia efektu grupowego myślenia, gdyż konformizm jest jednym z komponentów tego zjawiska [Aronson, 2004]. Wyższa wartość wskaźników:  $W_{14}$  i  $W_{15}$  świadczy o większej sile aktywacji efektu konformizmu. Globalna pozycja danej grupy FGI na wymiarze nasilenia konformizmu jest sumą 2 rang częściowych określających jej pozycję pod względem wielkości wskaźnika  $W_{14}$  i  $W_{15}$ , przy czym im wyższa wartość obu wskaźników, tym grupie przypisywano wyższą rangę. Globalny wynik określający nasilenie tego efektu zawiera się w przedziale 2–10, gdzie 10 oznacza najwyższą relatywną siłę aktywacji konformizmu.

Z kolei do pomiaru nasilenia syndromu grupowego myślenia zastosowano zestaw 5 wskaźników konkretyzujących przejawy działania tego zjawiska. Do tych symptomów zalicza się, np. zdominowanie dyskusji przez wąską podgrupę rozmówców, faworyzację orientacji na osiągnięcie stanu jednomyślności i utrzymanie bezkonfliktowej atmosfery nad różnorodnością opinii, powstrzymywanie się od ekspresji krytycznych uwag/wątpliwości na rzecz artykulacji reakcji potakujących. Wśród symptomów tego syndromu wskazuje się również tendencję do wyłączenia analitycznego stylu myślenia przejawiającą się w skracaniu czasu poświęcanego na przedyskutowanie atutów alternatywnych rozwiązań, identyfikację i korektę słabszych stron podjętego wyboru [Szmátka, 2007]. Inspirując się przesłankami zaczerpniętymi z teorii grupowego myślenia Janisa [Janis, 1972, 1982], skonstruowano 5 wskaźników:

1) wskaźnik monopolizacji dyskusji przez najbardziej aktywnego werbalnie rozmówcę ( $W_{10}$ ) – im wyższa wartość  $W_{10}$ , tym zakłada się wyższe nasilenie grupowego myślenia

$$W_{10} = \frac{t_d}{t_g} \cdot 100\% - \frac{100\%}{N}$$

gdzie:

$N$  – liczebność grupy,

$t_d$  – czas trwania wszystkich wypowiedzi wygenerowanych przez najbardziej aktywnego rozmówcę w 3 blokach tematycznych dyskusji utrzymanych w konwencji eksperymentalnej,

$t_g$  – czas wypowiedzi wszystkich dyskutantów (uwzględnienie tylko części eksperymentalnej);  $\frac{100\%}{N}$  – oczekiwany dla jednego rozmówcy udział procentowy jego wypowiedzi



– wskaźnik  $W_{10}$  informuje, o jaki procent „norma oczekiwanej aktywności” ( $100\%/N$ ) została przekroczona przez lidera aktywności werbalnej; innymi słowy, jaka jest nadwyżka czasu trwania jego wypowiedzi;

2) wskaźnik  $W_{11}$ , czyli liczba wypowiedzi, w których dyskutanci porównują dwie opcje lub eksponują słabsze strony wiążące się z wyborem danej opcji lub konstruktywnie wyrażają wątpliwości wobec argumentów rozmówców – im mniej wypowiedzi świadczących o aktywacji analitycznego trybu przetwarzania danych, tym zakłada się wyższe nasilenie myślenia grupowego;

3) wskaźnik  $W_{12}$ , czyli łączny czas trwania dyskusji pomniejszony o czas artykulacji wypowiedzi dygresyjnych – im krótszy czas dyskusji (wyrażony w sekundach), tym zakłada się wyższe nasilenie tego efektu;

4) wskaźnik  $W_{13}$ , czyli rozstęp liczby odpowiedzi – jest to różnica między liczbą odpowiedzi zaproponowanych przez lidera pomysłów, a najmniejszą liczbą odpowiedzi – im wyższa wartość  $W_{14}$ , tym oczekuje się bardziej nasilonej aktywacji syndromu grupowego myślenia;

5) wskaźnik  $W_{14}$ , czyli stosunek liczby wypowiedzi potakujących do opozycyjnych

– liczba wypowiedzi potakujących i opozycyjnych jest wyliczana dla całej dyskusji

– im wyższa wartość wskaźnika  $W_{14}$ , tym badani bardziej orientują się na podtrzymywanie życzliwej atmosfery spotkania niż krytyczne odnoszenie się do wypowiedzi rozmówców – tym samym zakłada się wyższe nasilenie grupowego myślenia.

Ustalono, iż dla wyższego nasilenia syndromu grupowego myślenia są diagnostyczne: wyższa wartość wskaźników:  $W_{10}$ ,  $W_{13}$  i  $W_{14}$  oraz niższa wartość wskaźników:  $W_{11}$  i  $W_{12}$ . Sumaryczny wynik wskazujący na siłę aktywacyjną tego efektu przyjmuje wartości z zakresu 5–25, gdzie 25 (iloczyn 5 wymiarów i maksymalnej wartości rangi równej 5) oznacza jego relatywnie najwyższe nasilenie.

Reasumując celem oszacowania relatywnego nasilenia efektu grupowego myślenia, proponujemy zestaw 5 następujących wskaźników: monopolizacja przestrzeni dyskusyjnej przez lidera aktywności werbalnej ( $W_{10}$ ), liczba wypowiedzi świadczących o analitycznym stylu przetwarzania informacji ( $W_{11}$ ), czas trwania dyskusji wymagającej porównania 2 opcji decyzyjnych ( $W_{12}$ ), rozstęp liczby odpowiedzi ( $W_{13}$ ) i stosunek częstości wypowiedzi potakujących do opozycyjnych ( $W_{14}$ ).

Z kolei celem oszacowania relatywnego nasilenia konformizmu sugerujemy zastosowanie 2 wskaźników, takich jak: stosunek wypowiedzi potakujących do opozycyjnych ( $W_{14}$ ) i siła tendencji do powstrzymywania się przed publiczną

artykulacją odmiennych przekonań, mimo dostrzegania poglądów odmiennych od własnych przekonań ( $W_{15}$  – WRO/WO).

Jak wspomniano, dla praktyków badań społecznych i marketingowych korzystających z metody zogniskowanego wywiadu grupowego kluczowym celem jest pozyskanie informacji umożliwiających zleceniodawcom badania rozstrzygnięcie pytań problemowych lub problemów decyzyjnych – informacji różnorodnych, unikalnych i nasyconych. Jak wykazały wyniki naszych badań, poszczególne wymiary metodologicznej jakości badania FGI, in. siła aktywacji efektów grupowych, wpływają na jakość informacji pozyskanych za pomocą tej metody (rozważania te zostaną podjęte w ostatniej części artykułu). Zatem warto ukierunkować przyszłe badania w stronę poszukiwania czynników potencjalnie różnicujących nasilenie pozytywnych i negatywnych efektów grupowych. Można na przykład sprawdzać, jak cechy osobowości i temperamentu, predyspozycje intelektualne, style przetwarzania informacji i kompetencje interpersonalne dyskusantów lub preferowany przez moderatora styl kierowania dyskusją wpływają na nasilenie efektów grupowych. Nasz artykuł prezentuje zestaw wskaźników i zasad (zalecenia konwencji eksperymentalnej) pozwalających w sposób zobiektywizowany (tj. minimalizujący wpływ odmiennej aktywności moderatora w różnych sesjach) zweryfikować, czy rozważana tu zmienna jest faktycznie czynnikiem różnicującym nasilenie efektów grupowych i w rezultacie pośrednio wpływającym na jakość pozyskanych informacji.

### **PROPOZYCJA OCENY JAKOŚCI PRAKTYCZNEJ FOKUSA**

W przypadku diagnozy praktycznej jakości badania FGI przyjętą przez nas jednostką analizy jest każda wypowiedź dyskusanta wygenerowana na przestrzeni całego badania i zgodna z intencją pytania moderatora. Uwzględniliśmy też wypowiedzi wygenerowane przez badanych w toku tradycyjnej konwencji fokusowej, ponieważ w naszej opinii różne wymiary praktycznej jakości fokusa (różnorodność, nasycenie i unikalność) w większym stopniu zależą od zdolności oraz preferencji poznawczo-intelektualnych badanych niż od przyjętego przez moderatora stylu kierowania dyskusją. Z analizy wyłączyliśmy wypowiedzi zgodne z problematyką dyskusji, ale niezgodne z intencją pytania moderatora. Dla przykładu, w toku dyskusji moderator zadaje pytanie – *co sądzicie na temat dodatkowych możliwości, jakie oferuje wam szkoła?*, po czym jeden z dyskusantów odpowiada – *uwagam, że w szkole jest za dużo zajęć i należałoby skrócić plan lekcji*. W takim przypadku pomimo, że wypowiedź dyskusanta dotyczy

szkolnictwa licealnego, to nie zostanie poddana analizie, gdyż nie jest ona zgodna z intencją pytania moderatora. W naszej opinii niezasadnym zabiegiem jest również tworzenie kategorii określających wizerunek szkoły na podstawie rekonstrukcji wizerunku uczniów. Na przykład na podstawie wypowiedzi *90% osób chodzących do liceum jest leniwych* jesteśmy w stanie utworzyć kategorię dotyczącą charakterystyki uczniów, ale już nie szkoły. Kategoria *instytucja pełna leniwych uczniów* nie dostarcza nam informacji o szkole, ale o uczniach. Bardziej wartościowa byłaby kategoria *szkoła jako instytucja nie potrafiąca motywować uczniów do nauki* – ale tworzenie takiej kategorii na podstawie przytoczonego materiału uznajemy za ryzykowną interpretację nie przystającą do danych – uczniowie mogą być leniwi również z powodu wpływów wywieranych przez kolegów, czy nieformalnych przepisów roli ucznia.

Powyższe przykłady wypowiedzi dyskutantów stanowią rodzaj odpowiedzi, których zakres tematyczny pozostaje w związku z problematyką badania, ale nie jest zgodny z intencją pytania moderatora. Co za tym idzie, są to wypowiedzi wyłączone z dalszej analizy, podobnie jak wątki dygresyjne, czyli wypowiedzi badanych, które oprócz tego, że nie są zgodne z intencją pytania moderatora, to nie pozostają w związku z problematyką dyskusji. Przykładem wątku dygresyjnego jest poniższa sekwencja wypowiedzi badanych, stanowiąca fragment transkrypcji:

7: *Ono [fiat seicento] potrafi wiele przetrwać, no. Podczas, gdy kiedyś jechaliśmy do Warszawy z chłopakami, to dwie sarny zabiliśmy, ale mimo tego jakoś tam dojechaliśmy.*

5: *To trzeba było przekleić taką sarenkę i odkreślić dwie kreski [śmiejch w grupie]*

8: *Będzie straszło następne sarny.*

3: *A na kolację gulasz z dziczyzny.*

7: *Nie, jechaliśmy dalej, a później do McDonalda [śmiejch w grupie].*

4: *Nie ma, jak McDonald, no nie.*

2: *Ciekawe swoją drogą, skąd oni mają zapasy w tym McDonaldzie?*

7: *No właśnie, może to wywożą do McDonalda.*

Ponadto na potrzeby projektu badawczego zdecydowaliśmy się na wyłączenie z analizy każdej wypowiedzi, której treść jest na tyle wieloznaczna lub niesprecyzowana, że groziłaby dokonaniem przez nas interpretacji nieadekwatnej do treści zebranych danych. Zatem analizie proponujemy poddawać tylko jednoznaczne wypowiedzi dyskutantów, które w swojej treści ściśle odpowiadają intencji pytania moderatora.

Celem diagnozy praktycznej jakości materiału zebranego z poszczególnych sesji FGI, należy dokonać operacjonalizacji „bogactwa informacyjnego”. W tym

celu skonstruowaliśmy 3 wskaźniki – wskaźnik zróżnicowania, nasycenia oraz unikalności. Poniżej prezentujemy opis każdego z nich.

Pierwszym wskaźnikiem niezbędnym do zdiagnozowania bogactwa informacyjnego uzyskanego materiału jest wskaźnik zróżnicowania (WZ), czyli liczba kategorii odpowiedzi wyodrębnionych przez badacza w oparciu o treść wypowiedzi wygenerowanych przez grupę FGI w następstwie pytania moderatora. W tym miejscu warto odwołać się do sposobu definiowania kategorii w ramach metodologii teorii ugruntowanej. W powyższej orientacji zakłada się, że „kategoria powstaje w procesie różnicowania elementów obserwowalnej rzeczywistości bądź zjawisk według jakiegoś kryterium” [Konecki, 2000: 29]. Dokonawszy przeniesienia powyższej definicji na potrzeby pracy uznajemy, iż kategorie odpowiedzi dyskutantów powstają poprzez różnicowanie ich wypowiedzi według kryterium zawartego w pytaniu moderatora. Przy tym podkreślenia wymaga fakt, że warunkiem wyodrębnienia danej kategorii jest istnienie argumentów, które ją uzasadniają. W naszej opinii taki, a nie inny sposób wyróżnienia poszczególnych kategorii odpowiedzi jest bardziej uzasadniony niż proste sumowanie odpowiedzi dyskutantów, prowadzące z reguły do mnożenia synonimicznych wypowiedzi nie zwiększających wiedzy badacza. Załóżmy, że na pytanie: *Jakim samochodem mogłaby być uczelnia?* dany dyskutant wskazuje zarzewiały polonez, argumentując swój wybór w następujący sposób: *Moja uczelnia to stara kamienica. Dlaczego zarzewiały? Bo moja uczelnia przecieka. I w niektórych salach nie można używać światła, bo grozi to spięciem*, a kolejny dodaje: *Wołga, ponieważ Wołga była z założenia takim samochodem, który miał być takim wschodnim mercedesem, takim ekskluzywnym, ale ciągle po takiej starej radzieckiej myśli technologicznej. A wiem, że studia są takie, że starają się być takie nowoczesne, starające się być takie dostosowane, podciągnięte pod Zachód, jeśli chodzi powiedzmy o taki sprzęt, wygląd tego wszystkiego, ale jednak ciągle gdzieś jest coś takiego, co nam się kojarzy z taką swojską, wschodnią częścią Europy*. Na podstawie tego materiału jesteśmy w stanie wyodrębnić dwie kategorie odpowiedzi rozumiane jako dwa różnego rodzaju znaczenia nadane uczelni na drodze metaforyzacji – porównania jej do samochodu. Pierwszą kategorią jest *szkoła jako budynek o niezadawalającym stanie technicznym*, a drugą jest *szkoła jako instytucja bezskutecznie aspirująca do zachodnich standardów*. Z kolei, jeśli w odpowiedzi na postawione wyżej pytanie dyskutanci wskazują poloneza, malucha i trabanta, za każdym razem uzasadniając swoje odpowiedzi w podobny sposób, to mimo tego, że wypowiedzi dyskutantów odwołują się do różnych modeli samochodów, na ich podstawie utworzymy tylko jedną kategorię. Taką decyzję motywujemy tym, że uzasadnienie wyboru każdego z wyżej wy-

mienionych samochodów odnosi się do równoważnych treściowo argumentów, czyli nie zwiększa bogactwa informacyjnego. Tym samym, im wyższa wartość WZ, tym większe zróżnicowanie odpowiedzi dyskutantów. Nie wyobrażamy sobie sytuacji, w której analiza materiału uzyskanego z wykorzystaniem metody jakościowej bazowałaby tylko na samych prostych operacjach arytmetycznych. Specyfika zogniskowanego wywiadu grupowego wymaga przyjęcia podejścia uwzględniającego przede wszystkim treść, a nie liczbę wypowiedzi dyskutantów.

Kolejny wskaźnik – wskaźnik nasycenia kategorii (WN) określa liczbę rozłącznych, tzn. zróżnicowanych treściowo argumentów (uzasadnień lub przykładów) przytoczonych przez badanych celem wyjaśnienia swoich wypowiedzi. Sparafrazowane uzasadnienia, niosące tą samą treść będą traktowane jako jeden argument uzasadniający konkretną kategorię.

Wskaźnik WN obliczamy według poniższego wzoru:

$$WN_x = \frac{nAK_1 + nAK_2 + nAK_k}{nK}$$

gdzie:

- x – numer grupy fokusowej,
- nAK<sub>1</sub> – łączna liczba argumentów uzasadniających pierwszą kategorię,
- nAK<sub>2</sub> – łączna liczba argumentów uzasadniających drugą kategorię,
- nAK<sub>k</sub> – łączna liczba argumentów uzasadniających n – tą kategorię,
- nK – łączna liczba kategorii wygenerowanych przez dyskutantów, a wyodrębnionych przez badacza.

Zatem wskaźnik WN jest ilorazem sumy łącznej liczby argumentów uzasadniających wszystkie wygenerowane przez badacza kategorie do łącznej liczby tych kategorii. Wskaźnik WN skonstruowaliśmy w oparciu o średnią arytmetyczną. Zabieg ten uznajemy za rodzaj uproszczenia, które posłuży nam do relatywizacji wyników otrzymanych dla poszczególnych sesji FGI. Nie zdecydowaliśmy się na zastosowanie średniej ważonej, ze względu na to, że dany argument może nie zwiększać bogactwa informacyjnego materiału, a mimo to inspiruje kolejną wypowiedź. Stąd nie jesteśmy uprawnieni do hierarchizacji znaczenia argumentów. Reasumując, za miarę nasycenia kategorii uznajemy różnorodność argumentów w postaci uzasadnień lub przykładów przytoczonych przez dyskutantów w nawiązaniu do określonej wypowiedzi. Im wyższa wartość wskaźnika WN, tym wyższy stopień nasycenia kategorii.

Z kolei wskaźnik unikalności (WU) definiujemy jako liczbę odpowiedzi unikatowych na pytanie moderatora. Przez określenie „unikatowa” mamy na

myśli taką wypowiedź dyskutanta, która we wszystkich sesjach pojawiła się tylko podczas jednego fokusa. Wskaźnik WU informuje o tym, na ile materiał uzyskany podczas danej dyskusji jest nieszablonowy, oryginalny i wychodzący ponad utarte schematy myślenia. Im wyższy wskaźnik WU, tym bardziej niekonwencjonalne spojrzenie na problem będący przedmiotem dyskusji.

Nie jesteśmy w stanie stwierdzić na ile pełne jest zróżnicowanie odpowiedzi dyskutantów lub nasycenie wybranej kategorii, czy też na ile materiał zebrany podczas dyskusji jest unikalny. Naszym celem jest relatywizacja uzyskanych wskaźników – porównywanie wielkości wskaźnika właściwego dla danej sesji ze wskaźnikami otrzymanymi dla pozostałych sesji. Innymi słowy, interpretacja tych 3 wskaźników pozwala wnioskować, czy stopień różnorodności, nasycenia lub unikalności kategorii wyodrębnionych przez badacza dla konkretnej sesji jest większy, mniejszy lub porównywalny z innymi sesjami.

Zaprezentowane wskaźniki sugerujemy wyliczać rozłącznie dla pytań zasadniczych oraz pytań projekcyjnych. Rozróżnienie to podyktowane jest odmienną charakterystyką pytań postawionych badanym. Pytania projekcyjne bardziej niż pytania zasadnicze angażują myślenie twórcze, którego istotą jest giętkość, plastyczność, zdolność do realizacji operacji intelektualnych na materiale abstrakcyjnym, wieloznacznym i słabo ustrukturalizowanym. Tym samym powyższe zadania mogą być zróżnicowane pod względem trudności i w efekcie mogą one różnicować stopień różnorodności i uzasadnienia wypowiedzi dyskutantów.

W celu zidentyfikowania sesji, która w porównaniu z pozostałymi pozwoliła na zebranie najbogatszego materiału przyjęliśmy założenie wskazujące, że im wyższe wskaźniki zróżnicowania, nasycenia oraz unikalności tym lepsza praktyczna jakość danego fokusa. Tym samym, uwzględniając przyjęte wskaźniki bogactwa informacyjnego, można sprecyzować, że im większa liczba kategorii, rozłącznych argumentów uzasadniających daną kategorię oraz wypowiedzi unikatowych, tym wyższa merytoryczna jakość fokusa.

### **ZAKOŃCZENIE – RELACJA MIĘDZY METODOLOGICZNĄ A PRAKTYCZNĄ JAKOŚCIĄ BADANIA FGI**

W końcowym fragmencie niniejszego artykułu pragniemy podjąć próbę wyjaśnienia relacji między metodologiczną a praktyczną jakością fokusa w świetle danych uzyskanych na drodze przeprowadzonych przez nas badań.

Po pierwsze stwierdzono, że wyższa siła aktywacji efektów emocjonalnych, tj. efektu bezpieczeństwa i spontaniczności może stymulować dyskutantów z grup

FGI do generowania bardziej różnorodnych lub unikalnych wypowiedzi. Emocje kontrolujące (np. lęk) zawężają zakres uwagi podmiotu do świata wewnętrznych przemyśleń, z kolei emocje adaptacyjne (np. ożywienie, zapał) ukierunkowują uwagę w stronę bodźców działających w otoczeniu zewnętrznym. Przenosząc to ogólne założenie psychologiczne na pole badań FGI można stwierdzić, że dyskutant przeżywający lęk przed, np. krytyką rozmówców lub odrzuceniem ze strony grupy może być nadmiernie zaabsorbowany antycypacją potencjalnych kosztów i strat wiążących się z wprowadzeniem do dyskusji nowego wątku tematycznego lub z publiczną artykulacją niekonwencjonalnych lub odmiennych od interlokutorów przekonań. Silniejsza tendencja do przeżywania emocji kontrolujących aktywowanych na myśl o przystąpieniu do badania może skłonić dyskutanta do rezygnacji z wygłoszenia własnej opinii lub zachęcić do dokonania treściowej replikacji poprzednich poglądów, które nie wzbudziły w grupie oporu – w ten sposób badany nie przyczynia się do wzrostu liczby wypowiedzi unikalnych i treściowo zróżnicowanych.

Po drugie, również wyższa siła aktywacyjna syndromu grupowego myślenia może powstrzymać przed artykulacją niekonwencjonalnych lub nowych w aspekcie treściowym wypowiedzi. Osiowym symptomem powyższego efektu jest dążenie do utrzymania życzliwej i bezkonfliktowej atmosfery oraz zachowania spójności grupowej, a każdy pogląd, zwłaszcza niekonwencjonalny generuje ryzyko ujawnienia się w grupie różnych stanowisk, mieszanych nastawień. W rezultacie taki pogląd może stać się przyczynkiem do zawiązania się konfliktu i rozpadu grupy.

Po trzecie, mniejsze nasilenie pozytywnych efektów zadaniowych, tj. efektu stymulacji i śnieżnej kuli może być barierą powstrzymującą przed pozyskaniem bardziej pogłębionych wypowiedzi, szczególnie gdy dyskutanci przejawiają słabszą tendencję do postrzegania rozmówców jako atrakcyjnych interpersonalnie. Im mniejsza skłonność do przypisywania członkom własnej grupy fokusowej atrybutów atrakcyjności interpersonalnej, np. inteligencji, tendencji do oryginalnego myślenia, umiejętności przekonującego argumentowania własnych poglądów, tym słabsza motywacja do intensywnego lub długotrwałego angażowania zasobów uwagi oraz pamięci długotrwałej umożliwiających odnoszenie się do spostrzeżeń innych na drodze komentowania, uzupełniania i klaryfikowania wypowiedzi innych i zadawania pytań.

Po czwarte, można zauważyć, że istnieje przełożenie między metodologiczną a praktyczną jakością badania FGI, dlatego wysiłki badawcze warto ukierunkować w stronę poszukiwania czynników regulujących oraz modyfikujących stopień nasilenia efektów grupowych. Wyniki naszego projektu badawczego mogą inicjować

ważny z praktycznego punktu widzenia trend w badaniach FGI – trend, którego celem jest identyfikacja regulatorów siły aktywacji efektów grupowych i tym samym pośrednio wpływających na różnorodność, unikalność i nasycenie informacji pozyskanych z wykorzystaniem metody zogniskowanego wywiadu grupowego.

## BIBLIOGRAFIA

- Akert R.M., Aronson E., Wilson T.D. [2004], *Psychologia społeczna*, Zysk i S-KA Wydawnictwo, Poznań.
- Bernacka R.E. [2004], *Konformizm i nonkonformizm a twórczość*, Wydawnictwo UMCS, Lublin.
- Domachowski D. [2007], *Przewodnik po psychologii społecznej*, Wydawnictwo Naukowe PWN, Warszawa.
- Janis Irving L. [1972], *Victims of Groupthink*, New York: Houghton Mifflin.
- Janis Irving L. [1982], *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Second Edition. New York: Houghton Mifflin.
- Konecki K. [2000], *Studia z metodologii badań jakościowych. Teoria ugruntowana*, Wydawnictwo Naukowe PWN, Warszawa
- Lisek-Michalska J. [2007], Lisek-Michalska J., Daniłowicz P., (red.) (2007) *Zogniskowany wywiad grupowy. Studia nad metodą*, Wydanie II poszerzone, Łódź: Wyd. UŁ
- Malinowski H. [2007], *Badanie dynamiki grupy w fokusie*, [w:] *Zogniskowany wywiad grupowy. Studia nad metodą* (red.) P. Daniłowicz, J. Lisek-Michalska, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Nikodem ska - Wołow nik A.M. [1999], *Jakościowe badania marketingowe*, Polskie Wydawnictwo Ekonomiczne
- Oyster C. K. [2002], *Grupy*, Wydawnictwo Zysk i S-ka, Poznań
- Szmatka J. [2007], *Małe struktury społeczne*, Wydawnictwo Naukowe PWN, Warszawa
- Tarkowska E. [1999], *Konformizm*, [w:] *Encyklopedia Socjologii*, t. IV., (red.) Z. Bokszański, Oficyna Naukowa, Warszawa.

Jolanta Lisek-Michalska  
Marcin Lipiec  
Anna Olczyk

## METHODOLOGICAL AND PRACTICAL EVALUATION OF FOKUS GROUP INTERVIEW (FGI) – A PROPOSAL OF EVALUATIVE PROCEDURE

### Abstract

The main goal of FGI praxis is to gather (diverse, rich and unique) pieces of information that can help clients, for whose companies the research is conducted, address the problems or decisions that they need to make. Based on our research results we can claim that the strength of the activation of different group effects influences the quality of merit in FGI research. To be precise, the dimension



of information richness is regulated by the strength of the activation of the stimulation effect and the snowballing effect. The other two dimensions (diversity and uniqueness) derive from the level of the intensity of the security effect, the spontaneity effect and the group thinking syndrome.

If there is a relation between methodological and practical research quality in FGI, it is worth looking into factors that can differentiate the strength of the activation of these group effects.

Our article is a proposal (of how it might be done) that involves indicators and rules allowing for the objective testing of whether an analyzed variable (e.g. a feature of character of a focus group participant) is an actual factor that differentiates the escalation of group effects and at the same time indirectly affects the practical quality of FGI.

**Key words:** focus group interview, group dynamics, quality of FGI research

## RECENZJA

SYLWIA MĘCFAL  
Uniwersytet Łódzki

**Klaus Krippendorff, Content Analysis. An Introduction to Its Methodology (Second Edition), Thousand Oaks, CA: Sage Publications, 2004.**

Książka Klause Krippendorffa<sup>1</sup> to druga edycja pracy poświęconej analizie zawartości. Krippendorff twierdzi, że od czasu publikacji pierwszej edycji (1980 rok) znacznie zmieniły się metody analizy zawartości oraz zmieniła się także specyfika tego, co poddawane jest analizie, a zatem specyfika materiałów, które są analizowane. W przedmowie do drugiej edycji książki autor wyraża pogląd, że w momencie opublikowania pierwszej wersji książki analiza zawartości znajdowała się na rozstaju dróg (*at a crossroads*; s. xiii). Badacze, stosujący tę technikę, mieli wówczas, zdaniem autora, wybór: albo kontynuować płytką grę polegającą na zliczaniu, za którą kryło się dziennikarskie zafascynowanie liczbami i bardzo wąskie pojęcie nauki, w którym tylko ilościowo mierzalne dowody się liczą, albo mogli przeorientować metody analizy zawartości w kierunku takich zjawisk społecznych, które są zarówno wywoływane przez teksty oraz obrazy, ale także są w nich zawarte, dlatego muszą być rozumiane poprzez ich tekstualne i wizualne elementy (s. xiii). Od tamtego momentu rzeczywistość przeszła

---

<sup>1</sup> Klaus Krippendorff jest profesorem komunikowania oraz profesorem cybernetyki, języka i kultury (the Gregory Bateson professor for Cybernetics, Language, and Culture) w Uniwersytecie Pensylwanii (University of Pennsylvania) w Szkole Komunikowania im. W. Annenberga (the Annenberg School of Communication). W 2004 roku autor otrzymał nagrodę Międzynarodowego Stowarzyszenia Komunikowania (International Communication Association) za recenzowaną przeze mnie książkę: „Content Analysis. An introduction to Its Methodology”. Zainteresowania profesora obejmują m.in. epistemologię konstruktywistyczną, cybernetykę drugiego rzędu, matematyczne podłoże cybernetyki, teorie systemowe, teorie komunikowania oraz informacji, metodologię nauk społecznych (zwłaszcza analizę zawartości) i inne.

ogromne zmiany i choć autor uważa, że metody analizy zawartości prezentowane w pierwszej edycji książki przetrwały wyzwania przed nimi postawione, to jednak tekstualna materia współczesnego społeczeństwa przeszła radykalną transformację, w dużej części poprzez ciągle trwającą rewolucję informacyjną i technologiczną. Rezultatem tych zmian jest, zdaniem autora, przeniesienie analizy zawartości, zwłaszcza tej wspomaganą technikami komputerowymi, w centralne miejsce metod, dzięki którym społeczeństwo poznaje siebie (s. xiv).

We wstępie autor wyróżnia trzy charakterystyczne cechy współczesnej analiza zawartości (s. xvii-xxi):

1) empiryczne ugruntowanie – analiza zawartości jest procesem eksploracyjnym oraz ma na celu wyjaśnianie i wyciąganie wniosków, 2) wykraczanie poza tradycyjne rozumienie symboli, zawartości oraz celowości – wskutek zmian w procesie komunikacji oraz dzięki postępowi technologicznemu badacze zajmujący się analizą tekstów nie mogą skupiać się na symbolach czy odwzorowaniach ani też nie mogą ograniczać się do pytań: kto mówi, co mówi, jakimi kanałami, do kogo i z jakim efektem<sup>2</sup>. Także uproszczone rozumienie pojęcia „zawartości” straciło swoją zdolność do wyjaśniania, 3) konieczność wytworzenia własnej metodologii, która pozwoli badaczom na planowanie, przeprowadzanie, komunikowanie, powtarzanie oraz krytyczne ocenianie prowadzonych analiz niezależnie od pojedynczych wyników (s. xx).

Książka składa się z trzech części, 14 rozdziałów: część I (3 rozdziały) – Konceptualizacja analizy zawartości, która zaczyna się krótkim rozdziałem dotyczącym historii metody, następnie autor zajmuje się zdefiniowaniem analizy zawartości w odróżnieniu do innych metod oraz sposobami zastosowania analizy zawartości; część II (6 rozdziałów) – Elementy analizy zawartości – która przedstawia procedury stosowane w analizie zawartości; część III (5 rozdziałów) – Analityczne drogi (*paths*) oraz techniki ewaluacyjne, gdzie autor przygląda się różnym sposobom radzenia sobie z analizą zawartości: od prezentowania danych, wyciągania wniosków z opracowywanych danych do używania technik komputerowych i ewaluacji dokonanych analiz za pomocą ustalenia ich trafności i rzetelności. Rozdział kończący tę część jest, jak mówi autor, praktycznym przewodnikiem, który podsumowuje toczące się dyskusje z punktu widzenia praktyka analizy zawartości. (s. xxii) Książka liczy 413 stron. Zakończona jest szeroką bibliografią, szczegółowym indeksem użytych pojęć i nazwisk oraz krótką informacją o autorze publikacji.

---

<sup>2</sup> „who says what, through which channel, to whom, and which effects” (Krippendorff, s. xix, za: Lasswell, 1960).

Autor zawiera też we wstępie „instrukcję obsługi” książki, sugerując od jakich rozdziałów powinno się zacząć lekturę w zależności od stopnia zaawansowania czytelnika. Zauważa, że książka powstała na podstawie jego doświadczeń podczas prowadzenia zajęć i seminariów dotyczących analizy zawartości i on sam myśli o tej książce jako o podręczniku dla zaawansowanych studentów studiów licencjackich lub początkujących studentów studiów magisterskich, choć może ona także służyć jako przewodnik dla praktyków oraz nauczycieli.

Każdy z rozdziałów książki omówię bardziej szczegółowo poniżej.

Rozdział 1 oferuje nam krótką historię analizy zawartości, zapoznając czytelnika z prekursorami omawianej metody, wczesnymi analizami zawartości: przez ilościowe analizy prasy (opisując zróżnicowane projekty badawcze od 1880 do początków XX wieku – gdy sam termin ‘analiza zawartości’ jeszcze nie istniał), przez wczesną analizę zawartości w latach 30. i 40. XX wieku do późniejszych analiz propagandy lat wojennych oraz rozwoju technik komputerowych analizy zawartości, a także podejść jakościowych (choć sam autor nie do końca zgadza się z podziałem analizy zawartości na ilościową i jakościową, gdyż sądzi, że „każde czytanie tekstu jest jakościowe, nawet jeśli pewne cechy charakterystyczne tekstu są później przekształcane w liczby” /s. 16<sup>3</sup>).

W rozdziale tym K. Krippendorff w sposób, w mojej ocenie, bardzo przystępny, pokazuje jak metoda ta, mająca korzenie w praktyce dziennikarskiej, a od 1941 roku znana jako analiza zawartości, zostaje przyswojona przez świat nauki: socjologów, badaczy mediów, psychologów, etnografów, politologów, lingwistów, literaturoznawców czy historyków. Ten ogromny wzrost zastosowania analizy zawartości spowodował, zdaniem autora, że ztracono sedno analizy zawartości, gdyż wszystko wydawało się możliwe do zbadania tą metodą, ale jednocześnie znacznie poszerzył możliwości metody, które pozwoliły uchwycić esencję ludzkich zachowań: mowę, rozmowę oraz komunikację pośrednią. Według autora obecnie obserwujemy ciągły wzrost możliwości zastosowania tej metody, w dużej mierze dzięki użyciu komputerów do przetwarzania materiałów oraz dzięki rosnącej liczbie materiałów, które są dostępne w formie elektronicznej.

W rozdziale 2 K. Krippendorff prezentuje metodologiczne i koncepcyjne podłoże analizy zawartości, a sam rozdział zaczyna od własnej definicji analizy zawartości: „Jest to technika badawcza używana do stawiania powtarzalnych oraz uzasadnionych (trafnych) wniosków z tekstów (czy też innej materii znaczącej)

---

<sup>3</sup> ‘Ultimately, All reading of texts is qualitative, even when certain characteristics of a text are later converted into numbers’

w odniesieniu do kontekstu, w którym te teksty są używane<sup>74</sup> (s. 18, tłumaczenie własne<sup>5</sup>).

Definicja ta zawiera w sobie zarówno pojęcia kluczowe dla innych metod badawczych: powtarzalność, trafność, jak i cechy charakterystyczne tylko dla analizy zawartości: pojęcie tekstu (nie ograniczające się do materiału pisanego, ale też odnoszące się do obrazów, dźwięków, map czy nawet liczb). Aby materiał został zakwalifikowany jako nadający się do analizy zawartości, musi mówić do kogoś o zjawiskach, które mogą być odczuwane lub obserwowane poza tym materiałem.<sup>6</sup> (s. 19, tłumaczenie własne). Autor wyróżnia w tym rozdziale także 3 typy definicji odnoszące się do analizy zawartości: 1. Definicje, które uważają ‘zawartość’ za coś stałego w tekście (tu autor zalicza między innymi klasyczne definicje Berelsona, Gerbnera); 2. Definicje, które uważają ‘zawartość’ za własność źródła danego tekstu (definicje Holstiego, Lasswella – koncentrujące się na intencjach nadawcy: co mówi?, jak mówi?, do kogo?, dlaczego? kto mówi? z jakim skutkiem?); [zob. Ogryzko-Wiewiórowski, 1984 Goban-Klas, 1997]; 3. Definicje, które uważają, że ‘zawartość’ wyłania się w procesie analizy tekstu i jest relatywna do określonego kontekstu (tu zalicza swoją definicję, od której zaczyna się ten rozdział). Odnosząc się do tej klasyfikacji w dalszej części rozdziału autor argumentuje, dlaczego opowiada się za trzecim typem definicji, podaje przykłady jej zastosowania, a także prezentuje schemat analizy zawartości (*framework*), na który składają się: tekst, pytania badawcze (problemowe), kontekst analizy, konstrukt analityczny, wnioski oraz sprawdzanie wiarygodności analizy. Rozdział kończy się porównaniem analizy zawartości do innych metod badawczych z intencją podkreślenia zalet omawianej metody.

Mimo że rozdział omawia trudną tematykę, napisany jest zrozumiałym i łatwo przyswajalnym językiem, a wprowadzane nowe terminy są przez autora skrupulatnie wyjaśniane, także poprzez podawanie przykładów. W tym miejscu bardzo przydałoby się przełożenie tych założeń na praktykę i pokazanie czytelnikowi zastosowania tych procedur, jednak tego tutaj nie znajdujemy.

Zamiast tego autor w rozdziale 3, który kończy pierwszą część książki, omawia przykłady przeprowadzonych już analiz. Pokazuje, w jakich celach analiza

<sup>4</sup> ‘Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use.’

<sup>5</sup> W książce „Mass media. Metody badań” w rozdziale poświęconym analizie treści R.D. Wimmer oraz J.R. Dominick także przytaczają definicję K. Krippendorffa (z pierwszej edycji książki): Jest to technika badań, „której celem jest ustalenie powtarzalnych i trafnych związków między danymi a ich kontekstem”, [2008, s. 211];

<sup>6</sup> ‘[...] they speak to someone about phenomena outside of what can be sensed or observed.’

zawartości bywa stosowana. On sam wyróżnia 6 kategorii analizy zawartości, które odnoszą się głównie do celów, jakie stawia przed sobą badacz posługujący się tą metodą: a) przewidywania, b) porównania do standardów, c) wskaźniki szerszych zjawisk, d) lingwistyczne reprezentacje, e) analiza konwersacyjna, f) procesy instytucjonalne (dodatkowo dwie pierwsze kategorie dzielą się jeszcze na podkategorie, czyli przewidywania dzielą się na trendy, wzory oraz różnice, a standardy na identyfikacje, ewaluacje oraz sądy). Konkludując rozdział autor prezentuje warunki owocnej analizy zawartości, a zatem: a) bliskość tematyki analizy do sposobów użycia języka, b) skupienie się na faktach ukonstytuowanych w języku, w sposobie użycia analizowanych tekstów (w tym na atrybucjach, społecznych powiązaniach użycia języka, na zachowaniach publicznych czy rzeczywistości instytucjonalnej). Zatem powtarzalne, rutynowe, publiczne oraz zinstytucjonalizowane zjawiska są bardziej przystępne do analizy niż te, które pojawiają się rzadko i są niekonwencjonalne (s. 77). Powodzenie analizy zawartości zależy także od badacza: im większy zasób słownictwa badacza, a także im bardziej świadomy jest on subtelnych konwencji językowych, także własnych, tym większe ma szanse na sukces podjętego przedsięwzięcia badawczego (s. 77).

Część druga publikacji ma na celu szczegółowe omówienie elementów analizy zawartości. Krippendorff zaczyna ją od rozdziału (4) zapoznającego odbiorcę z logiką tworzenia projektu badawczego opartego na analizie zawartości. Autor wymienia w tym rozdziale elementy analizy zawartości i krótko je charakteryzuje. Każdy z elementów jest następnie szczegółowo omówiony w kolejnych rozdziałach książki. Wyróżnione elementy analizy zawartości to: 1. Wybór jednostek analizy */unitizing/* (któremu poświęcony jest rozdział 5), 2. Dobór próby */sampling/* (rozdział 6), 3. Kodowanie materiałów */recording/coding/* (rozdział 7), 4. Redukowanie */reducing/* danych (statystyczne opracowywanie danych, upraszczanie czy też synteza danych, rozdział 8), 5. Wyciąganie wniosków o zjawisku */abductively inferring contextual phenomena/* (rozdział 9), 6. Sprawozdanie z uzyskanych odpowiedzi na postawione wcześniej pytania badawcze */narrating the answer to the research question/*. Rozdział 4 prezentuje także kilka wzorów postępowania badawczego */research designs/*, które mogą być zastosowane przy posługiwaniu się metodą analizy zawartości.

W rozdziale 5 dowiadujemy się, że podczas przeprowadzania analizy zawartości mamy do czynienia z różnymi 'jednostkami' */units/*: jednostkami doboru, jednostkami analizy (*coding/recording units*), a także z jednostkami kontekstu. Jednostki doboru wyznaczają i ograniczają badaczowi materiał do analizy, jednostki analizy wskazują te elementy tekstu, które mają być szczegółowo opisane i skategoryzowane, a jednostki kontekstu to takie elementy analizowanego tekstu,

które brane są pod uwagę przy opisie jednostek analizy<sup>7</sup>. Autor podkreśla dużą rolę wyróżnienia jednostek poszczególnych typów w prawidłowym określeniu, jakie dane mają zostać zaobserwowane i zakodowane oraz pokazuje, w jaki sposób możemy wyróżniać takie jednostki (według kryteriów fizycznych, syntaktycznych, przynależności do określonej kategorii, tematycznych, zawierających określone relacje między słowami, częściami zdania /*propositional distinctions*/).

Metody doboru materiałów do analizy są szczegółowo omawiane w rozdziale 6. Zawartość tego rozdziału będzie zapewne dobrze znana badaczom mającym doświadczenie z badaniami survey'owymi. Jednak autor ciekawie przedstawia znane metody doboru próby, przenosząc je na grunt analizy zawartości oraz podając jednocześnie przykłady zastosowania określonych metod doboru. Badaczom oraz studentom, którzy mają do czynienia z obszernymi materiałami do analizy, rozdział ten będzie niewątpliwie bardzo przydatny i pomoże w sposób uzasadniony utworzyć korpus materiałów przeznaczonych do analizy.

Rozdział 7 podkreśla wagę procesu kodowania materiałów, a także przypomina o utrwalaniu (nagrywaniu) materiałów, które chcemy poddawać analizie. Autor wskazuje również na istotne cechy, jakie powinien posiadać koder oraz na jego przeszkolenie. Duża część rozdziału poświęcona jest sposobom transkrypcji i zapisu danych do późniejszej analizy oraz przykładowym schematom kodowania danych.

Rozdział 8 nosi tytuł „Języki danych”. Tym barwnym tytułem autor daje do zrozumienia, że każdy zbiór danych zorganizowany w odpowiedni sposób ma swój język (a zatem swoją składnię oraz semantykę). Język danych opisuje, jak wszystkie kategorie, zmienne, notatki, transkrypcje czy dane elektroniczne łączą się ze sobą w jeden system (s. 150)<sup>8</sup>. Dane przemawiają do badaczy za pomocą zmiennych i ich wartości, stałych, a także określonych uporządkowań wartości zmiennych (np. w łańcuchy /*chains*/ – które przypominają skale, pętle czy koła /*tw. recursions* – czyli cyrkularne połączenia między wartościami zmiennych/, sześciiany /*cubes*/ czy drzewa /*trees*/). W rozdziale omawiane są także rodzaje zmiennych oraz związane z nimi dozwolone operacje matematyczne.

Część drugą książki kończy rozdział 9, w którym stajemy przed rozważaniami na temat wniosków, jakie możemy wyciągać na podstawie przeprowadzonej analizy zawartości, oraz szczególnej roli konstruktów analitycznych w tym

<sup>7</sup> Zatem jest to jakiś szerszy fragment analizowanego tekstu, który pozwala nam na odpowiednie zaklasyfikowanie jednostki analizy, np. cała wypowiedź, w której pojawiają się analizowane postaci w określonych rolach.

<sup>8</sup> ‘...a data language describes how all categories, variables, notations, formal transcripts, and computer-readable accounts hang together to form one system’

procesie. Razem z autorem zastanawiamy się także, jakie są źródła pewności i niepewności naszych wniosków.

Część 3 pracy, złożona z 5 rozdziałów, skupia się na tym, aby pomóc badaczom ocenić jakość przeprowadzanych analiz zawartości, a także pokazuje różne sposoby analizy i prezentacji danych uzyskanych tą metodą. Rozdział 10 pokazuje badaczom najbardziej powszechne sposoby analizy danych uzyskanych za pomocą analizy zawartości, a także możliwości prezentacji tych danych. Wspomniane są takie sposoby prezentacji i analizy danych jak: zestawienia tabelaryczne */tabulations/*, związane z prezentacją częstości występowania zjawisk (także zestawienia w postaci wykresów, czy analizy wzdłuż czasowe – analizy trendu); tabele krzyżowe związane z prezentacją związków między zmiennymi; techniki wielowymiarowe (np. analiza regresji, analiza ścieżek); analiza czynnikowa oraz wielowymiarowe skalowanie; analiza skupień */clustering/*; analiza semantyczna (sieci semantyczne) */semantic nodes/*, a także tworzenie wizerunków, portretów czy profili (za pomocą analizy atrybucji czy też specjalnych programów komputerowych). Część prezentowanych w tym rozdziale sposobów analizy będzie na pewno znana badaczom/analitykom, zwłaszcza tym zajmującym się danymi ilościowymi. Niektóre rozwiązania przedstawione czytelnikowi są jednak specyficzne dla analizy zawartości i mogą przydać się w praktyce, ułatwiając nie tylko analizę, ale też odbiór wyników analizy zawartości.

Bardzo dużo miejsca poświęca autor na rozważenie kwestii rzetelności (rozdział 11) oraz trafności (rozdział 13) prowadzonych analiz. Rozdział 11 jest bardzo obszerny. Kwestia rzetelności danych poruszana jest tutaj nie tylko w kontekście analizy zawartości, ale także badań naukowych w ogóle. Autor opisuje typy rzetelności (stabilność, reproduktywność */powtarzalność/* oraz dokładność), warunki uzyskiwania rzetelnych danych, a także bardzo szczegółowo odnosi się do sposobów pomiaru zgody koderów */coefficients agreement for coding/*, w tym do współczynnika  $\alpha$ -Krippendorffa, a także współczynników zgody wyboru jednostek analizy */ $\alpha$ -agreement for unitizing/*. Podobnie szczegółowo omawiana jest w rozdziale 13 typologia sposobów ustalania trafności.

Rozdziały te rozdzielone są najdłuższym fragmentem publikacji (rozdziałem 12), który dotyczy tego, jak komputery mogą wesprzeć prowadzone analizy zawartości. To głównie o ten rozdział została wzbogacona druga edycja książki K. Krippendorffa. Zaprezentowane są w nim nie tylko analityczne sposoby wykorzystywania komputerów, ale także wskazane są elektroniczne zbiory tekstów, w których można szukać materiałów źródłowych oraz programy mające wspomagać badaczy prowadzących analizy zawartości. Autor określa te ostatnie jako *'computer-aided text analysis (CATA) software'* podkreślając ich pomocniczą



rolę przy prowadzeniu analiz i przestrzegając przed nadmiernym optymizmem, który wzbudzą producenci różnorodnych programów. Rozdział ten nie tylko prezentuje najbardziej powszechne oprogramowania, ale także wskazuje na ich wady i zalety oraz omawia przykładowe zastosowania. Ten fragment książki jest na pewno bardzo ciekawy, zmusza czytelnika do zgłębiania wiedzy odnośnie omawianego oprogramowania oraz będzie na pewno przydatny dla badaczy chcących prowadzić analizę zawartości z wykorzystaniem technik komputerowych. Mam jednak obawy, że część uwag w nim zawartych nie będzie przydatna dla polskiego czytelnika, z racji na niedostępność niektórych wymienianych baz danych czy części oprogramowania.

Książkę kończy „Praktyczny przewodnik po analizie zawartości” (rozdział 14), gdzie krok po kroku śledzimy proces analizy zawartości od koncepcji badawczej, przez stworzenie projektu badawczego (*research proposal*), zastosowanie wcześniej napisanej koncepcji (czyli po prostu przeprowadzenie analizy), aż do zaprezentowania wyników. Pełni on rolę dłuższego podsumowania całego tekstu, choć koncentruje się na praktycznych uwagach dotyczących prowadzenia własnych analiz (wskazuje także na pułapki, w które można wpaść nie będąc doświadczonym badaczem).

Po książkę Klausa Krippendorffa sięgnęłam z kilku powodów: z ciekawości, gdyż na listach lektur uczelni brytyjskich w zakresie metodologii badań wymieniana jest jako klasyka; chciałam wzbogacić swoją wiedzę i warsztat w zakresie analizy zawartości, aby bardziej skutecznie prowadzić własne badania; chciałam wzbogacić swoją wiedzę w zakresie tej metody, aby wykorzystać ją do prowadzenia zajęć ze studentami; a także według mnie na polskim rynku brakuje zwartej publikacji dotyczącej analizy zawartości. Lektura nie do końca spełniła moje oczekiwania (zwłaszcza w zakresie praktyki badawczej), ale przyznam, że na pewno będę do niej wracać, gdyż podsunęła mi kilka pomysłów, uzupełniła braki w wiedzy, a także uporządkowała wiedzę już posiadaną.

Autor sam charakteryzuje ją jako podręcznik przydatny dla zaawansowanych studentów studiów I stopnia lub początkujących studentów studiów II stopnia. Według mnie jest to tekst za trudny dla studentów studiów licencjackich (choć polecam go dla ambitnych), ale na pewno przydatny dla przyszłych magistrów czy dla doktorantów, którzy zamierzają przeprowadzać analizę zawartości. Elementy tego tekstu powinny być włączone do programu metod badań społecznych, czy też powinien on być przynajmniej polecany jako lektura uzupełniająca. Niezbędną umiejętnością przy przyswajaniu tego obszernego materiału będzie na pewno dobra znajomość języka angielskiego. Niektóre z rozdziałów pod względem językowym stanowią prawdziwe wyzwanie.

Badacze bardziej wtajemniczeni w arkany analizy zawartości skorzystają na lekturze tej pozycji, gdyż jest to uporządkowany i szczegółowy wywód prowadzony przez eksperta w tej tematyce.

Wśród wielu pozytywnych aspektów książki, można dostrzec też pewne jej uchybienia. Moim zdaniem, mimo wielu przykładów przeprowadzonych analiz oraz praktycznego przewodnika na zakończenie, książce brakuje jednak strony praktycznej. Brak propozycji prostych ćwiczeń dla studentów, które oparte o rzeczywiste dane, mogłyby krok po kroku pokazać, jak przeprowadzić takie analizy (książka przecież aspiruje do roli podręcznika).

Poza tym, choć autor niechętnie używa podziału na ilościową i jakościową analizę zawartości, kiedy to czyni, można dostrzec, że „jakościowe” podejście (które autor rozumie jako mające korzenie w etnografii, antropologii kulturowej oraz perspektywie interpretatywnej, s. 303) jest w tym tekście trochę zaniedbane. Wspomniane jest ono w krótkich fragmentach tekstu w rozdziale 1, 4 oraz 12 (kiedy autor omawia programy komputerowe wspomagające jakościową analizę zawartości). Krippendorff zaznacza, że badacze jakościowi stosują inne kryteria (niż rzetelność i trafność) do oceny wyników prowadzonych analiz tekstu (s. 88–89). Za Denzinem i Lincoln (2000, polskie wydanie 2009) przytacza między innymi: pewność /*trustworthiness*/, wiarygodność /*credibility*/, możliwość przenoszenia [wyników] /*transferability*/, zakorzenianie rozumienia w doświadczeniu /*embodiment*/, odpowiedzialność /*accountability*/, refleksyjność /*reflexivity*/ oraz cele emancypacyjne /*emancipatory aims*/ . Jednak autor podkreśla także, iż jakościowe podejścia do analizy tekstu nie powinny być traktowane jako niezgodne z analizą zawartości. Schemat analizy zawartości [„ilościowej”], zdaniem autora, da się zauważyć także w podejściu jakościowym, choć ten pierwszy charakteryzuje się znacznie bardziej rozbudowaną fazą projektowania badania /*design phase*/ (s. 89). Badacze prowadzący analizę zawartości mogą także zaadoptować od badaczy jakościowych wykorzystywanie wielu kontekstów interpretacyjnych, wielość pytań badawczych czy większe refleksyjne zaangażowanie badacza. Jednak odniesienia do popularnych w Polsce badaczy zajmujących się jakościowymi analizami tekstu są w tej publikacji znikome. Autor wspomina znaną w Polsce publikację N.K. Denzina oraz Y.S. Lincoln „Metody badań jakościowych”, przytacza przykłady analiz prowadzonych przez badaczy dyskursu jak Van Dijk czy analizy konwersacyjne – Sacks, jednak są to jedynie wzmianki. Nie pojawiają się w tej książce w ogóle odniesienia do dość popularnych w Polsce prac Davida Silvermana.

Największym walorem publikacji jest według mnie refleksja metodologiczna. Badacze czasami nie myślą o analizie zawartości jako o metodzie, która też powinna spełniać pewne kryteria, powinna być rzetelna i trafna. Krippendorff

wielokrotnie podkreśla wagę dbania o „naukowość” analiz oraz sprzeciwia się traktowaniu analizy zawartości jako czynności, którą każdy wykonuje, czytając na co dzień różne teksty. Wartością tej pozycji jest też szereg przykładów analiz wykonanych przez badaczy z różnych dziedzin, dlatego może być ona źródłem wiedzy nie tylko dla badaczy/studentów z obszaru nauk społecznych czy psychologii, ale również dla tych związanych z marketingiem, zarządzaniem czy badaniami rynku.

#### **BIBLIOGRAFIA**

- T. G o b a n - K l a s, *Analiza zawartości przekazów prasowych*, [w:] Marian Malikowski i Marian Niezgoda (red.). *Badania empiryczne w socjologii. Wybór tekstów*, Tom 2, Tyczyn 1997: WSSG, ss. 294–320.
- H. O g r y z k o - W i e w i ó r o w s k i, *Zastosowanie analizy treści w badaniach socjologicznych*, *Annales Universitatis Mariae Curie-Skłodowska*, Vol. IX, 1, Lublin 1984, s. 1–13.
- R.D. W i m m e r, J.R. D o m i n i c k, *Mass media. Metody badań*, rozdział 6, s. 209–249, Wyd. Uniwersytetu Jagiellońskiego, Kraków 2008.

## ARTYKULY

- Anna Kubiak, Aneta Krzewińska – Sondaż deliberatywny – inwentarz problemów
- Franciszek Sztabiński, Teresa Żmijewska - Jędrzejczyk – *Mixed Mode Survey Design*: problem efektu techniki
- Rafał Boguszeński, Natalia Hipsz – Od kwestionariusza papierowego do wspomaganego komputerowo. Analiza porównawcza technik PAPI i CAPI
- Katarzyna Grzeszkiewicz-Radulska – Jak w Polsce rozwija się zjawisko niedostępności respondentów? Analiza na przykładzie badań CBOS 1993–2011
- Natalia Hipsz – Standaryzowane techniki badawcze jako szczególny sposób komunikowania się
- Wojciech Jabłoński – Sytuacje trudne w wywiadzie telefonicznym
- Sylwia Męćfał – Problemy badań terenowych – wybrane kwestie metodologiczne, praktyczne oraz etyczne przy badaniu zjawisk „trudnych”
- Rafał Maciąg – Społeczna konstrukcja dokumentów i danych urzędowych. Dane medyczne
- Jolanta Lisek-Michałska – Ocena metodologicznej i praktycznej jakości fokusa – propozycja procedury

## RECENZJA

- Klaus Krippendorff, *Content Analysis. An Introduction to Its Methodology* – rec. Sylwia Męćfał