

Katarzyna Grzeszkiewicz-Radulska*

METODY BADAŃ PILOTAŻOWYCH

Streszczenie. Tradycje polskich badań naukowych nad pilotażem związane są głównie z „łódzką szkołą metodologiczną” i sięgają lat 70. oraz 80. ubiegłego wieku. Obecnie zainteresowanie problematyką testowania i oceny narzędzi badawczych obserwuje się przede wszystkim na gruncie amerykańskim. Artykuł prezentuje dokonania w tej dziedzinie i przedstawia najważniejsze terenowe metody badań pilotażowych. Charakterystyka każdej z metod zawiera omówienie procedur badawczych, z uwzględnieniem sposobu doboru próby i opracowania materiałów, a także uwagi dotyczące praktycznych możliwości ich wykorzystania w terenie. Autorka dokonuje analizy wykorzystywanych w poszczególnych metodach źródeł informacji, omawia problemy dotyczące symulowania w pilotażu warunków charakterystycznych dla wywiadu właściwego oraz dokonuje podsumowania przedstawionych metod pod kątem stopnia, w jakim realizują cele postawione przed badaniami pilotażowymi.

Słowa kluczowe: pilotaż konwencjonalny, pilotaż pogłębiany, wywiad kognitywny, *behavior coding*, *response latency*, winiety, *respondent debriefing*, eksperyment.

1. Wprowadzenie

Badanie pilotażowe jest rodzajem badania wstępnego i ma na celu przetestowanie oraz ocenę narzędzia, które ma być wykorzystane w badaniu zasadniczym, realizowanym za pomocą wybranej techniki kwestionariuszowej. Mówiąc ogólnie, narzędzie sprawdzane jest pod kątem jego przydatności do dostarczania informacji poszukiwanych przez badacza. Analizując różne definicje pilotażu, nie zawsze *explicite* wyłożone, stwierdzamy, że pilotaż rzadko kiedy obejmuje ocenę ustaleń koncepcyjnych, np. trafności doboru wskaźników, choć decyzje w tej kwestii znacząco rzutują na przydatność narzędzia w osiągnięciu postawionych celów. W pilotażu sprawdza się najczęściej, czy zaproponowany sposób pomiaru zjawiska, czyli sformułowane pytanie kwestionariuszowe, tak a nie inaczej ułożone, mające określone brzmienie i strukturę, przynosi informacje, które mogą się stać podstawą uzasadnionych wniosków. Powody, dla których pytanie nie przynosi godnego

* Katarzyna Grzeszkiewicz-Radulska – Katedra Metod i Technik Badań Społecznych, Instytut Socjologii, Wydział Ekonomiczno-Socjologiczny, Uniwersytet Łódzki, ul. Rewolucji 1905 r. nr 41, 90-214 Łódź.

zauwania materiału, i w tym sensie jest wadliwe, mogą być rozmaite, np. występowanie w pytaniu błędów logicznych czy terminów niejednoznacznie rozumianych, zbytnia drażliwość pytania itd. Byłoby idealnie, gdyby pilotaż pozwalał trafnie wskazywać wady pytania, a także dawał wskazówki jak je przeredagować, aby było miarodajne, i w końcu dostarczał narzędzi pozwalających to stwierdzić.

Metodologowie zajmujący się problematyką pilotażu zauważają pewien paradoks – z jednej strony potrzebę sprawdzenia narzędzia podkreśla się już w tekstach dydaktycznych, a wagę tego postulatu zdaje się dostrzegać każdy badacz sondażowy; z drugiej strony te same podręczniki niewiele mówią, w jaki sposób to robić¹, a w empirycznych publikacjach naukowych rzadkością jest odnaleźć informację mówiącą o tym, czy pilotaż został przeprowadzony, za pomocą jakich metod i z jakim skutkiem (Presser i in. 2004: 109). Jest to podstawą domniemania, że w dużej części przypadków przeprowadzanie pilotażu jest albo zaniedbywane, albo przebiega według mało systematycznych reguł. Jak zauważył m.in. Oksenberg, „mimo iż od jakości narzędzia zależy wartość otrzymanych rezultatów, pilotaż to w praktyce najmniej naukowy aspekt badania sondażowego” (Oksenberg i in. 1991: 362).

W polskim piśmiennictwie naukowym zainteresowanie problematyką testowania i oceny narzędzia badawczego obserwuje się głównie na przestrzeni lat 70. XX w. i w początkach kolejnej dekady. Znaczący udział w tym mieli metodologowie pracujący w Instytucie Socjologii UŁ oraz IFiS PAN w Łodzi², w tym w szczególności profesor Krystyna Lutyńska, która opracowała koncepcję pilotażu pogłębionego (1974, 1975, 1984).

W Stanach Zjednoczonych ożywienie zainteresowania problematyką badań pilotażowych nastąpiło nieco później (w latach 80. XX w.), ale trwa do dziś. Zaowocowało ono wielością publikacji, w tym książkowych, które unaocniają duży rozwój w tej dziedzinie – opracowuje się nowe techniki oraz bada ich możliwości.

Zadaniem artykułu jest przybliżenie różnych rozwiązań stosowanych w badaniach pilotażowych, zarówno przez polskich, jak i zagranicznych badaczy. Jakkolwiek publikacja ta przedstawia je tylko w pewnym zarysie i ma charakter przeglądu literaturowego, przygotowana została z intencją (ponownego?) zainteresowania polskich badaczy i metodologów problematyką pilotażu, co biorąc pod uwagę szerokie i częste wykorzystanie badań sondażowych, pozostaje sprawą bardzo ważną.

Na początek ściślej określę zakres podejmowanych w artykule zagadnień. Choć w pilotażu wypróbowywane jest całe narzędzie, przedmiotem badania są przede wszystkim pytania kwestionariusza i na ten aspekt skierowana jest również moja uwaga w tym artykule. Jednocześnie przedstawione tu rozważania dotyczą

¹ Przykładem jest podręcznik E. Babbiego, *Badania społeczne w praktyce*, PWN, Warszawa 2003, s. 281–283.

² Efektem ich prac są liczne publikacje wydane w serii „Analizy i próby technik badawczych w socjologii” (szczególnie t. V) i „Z metodologii i metodyki socjologicznych badań terenowych” (szczególnie z. 3, 5, 7).

sposobów pilotowania narzędzia przygotowywanego do badań wykorzystujących technikę wywiadu, a nie ankiety. Mowa jest więc o technikach, które zakładają obecność ankietera jako osoby odczytującej i zapisującej odpowiedzi respondenta, a nie o tych, w których respondent sam wypełnia kwestionariusz. Nie omawiam problemów związanych z pilotowaniem kwestionariusza ani od strony rozwiązań graficznych, ani technicznych, które dotyczą narzędzi mających postać elektroniczną. Zawartym w tytule określeniem „metody” posługuję się w szerokim znaczeniu tego słowa i obejmuję nim stosowane praktyki, techniki badawcze i metody *sensu stricto*. Spośród tak rozumianych metod omawiam tylko te z nich, które zakładają przetestowanie i ocenę pytań kwestionariuszowych na drodze badań terenowych. Do grupy tej nie będą należeć na przykład konsultacje metodologiczne (brak empirycznych badań) czy zogniskowany wywiad grupowy stosowany z zamiarem dookreślenia problematyki i ułożenia pierwszej wersji kwestionariusza (brak narzędzia do testowania). Celem artykułu jest dostarczenie charakterystyki i porównanie różnych metod od strony wykorzystywanych źródeł informacji, procedur badawczych, rodzaju przynoszonych rezultatów i ich przydatności w przygotowaniu miarodajnego narzędzia. Prezentacja ta obejmuje: pilotaż konwencjonalny, pilotaż pogłębiony, wywiady kognitywne, różne techniki obserwacji, winiety, *respondent debriefing* oraz eksperymenty.

2. Pilotaż konwencjonalny

Termin „pilotaż konwencjonalny” odnosi się do dominującej (przynajmniej do niedawna) praktyki amerykańskich badań mających na celu przetestowanie narzędzia. Konwencjonalnie przeprowadzony pilotaż jest badaniem, w którym przestrzega się reguł mających obowiązywać w badaniu zasadniczym, z tym że realizowanym na nielosowej próbie, liczącej od kilku do kilkudziesięciu respondentów. Po przejściu takiej terenowej „próby generalnej”, od ankieterów oczekuje się ustnych bądź pisemnych raportów, w których zdadzą relacje z tego, jakie problemy respondentom lub im samym nastręczały pytania kwestionariusza (Presser i in. 2004: 110). Dostarczony w latach 80. przez K. Lutyńską opis standardowo przeprowadzonego pilotażu w Polsce wskazuje na zasadnicze podobieństwo postępowania (1984: 59).

Powszechność tego rozwiązania wynika, jak można sądzić, głównie z tego, że realizuje ono pewien program minimalny. Jak się jednocześnie okazuje, jeszcze w nieodległej przeszłości, przez kilku czołowych amerykańskich metodologów był on uznawany za wystarczający. Powodzenie w realizacji celu uzależniano przy tym często od umiejętności i kwalifikacji ankieterów (Presser i in. 2004: 110).

Podkreślane, zarówno wcześniej jak i obecnie, wady pilotażu konwencjonalnego koncentrują się na kilku kwestiach. Po pierwsze, w sytuacjach, w których odtwarza się warunki zasadniczego badania, czyli warunki standaryzowanego

wywiadu, pewne problemy, mimo że istnieją, nie zostaną ujawnione. Dotyczy to głównie przypadków, w których respondent inaczej niż badacz rozumie intencję pytania lub termin w nim zawarty, a przy tym odpowiedzi dostarcza płynnie, bez zauważalnych oznak np. zawahania czy prośby o wyjaśnienie, które nasunęłyby ankieterowi podejrzenie, że odpowiedź jest nietrafna. Jak często powtarzają metodologowie, fakt, że respondent udziela odpowiedzi, nie świadczy o tym, że zrozumiał pytanie w sposób, na jakim zależało badaczowi. Wagę tego stwierdzenia mogą unaocznic chociażby wyniki opracowania Nuckolsa (1953), który sprawdził sposób rozumienia 10 pytań szeroko stosowanych w ogólnokrajowych sondażach amerykańskich i ustalił, że każde z nich „generowało” niezgodności w rozumieniu – średnio 17% respondentów pojmowało pytanie inaczej niż badacz (cyt. za Foddy 2002: 328). Tymczasem, jak pokazują wyniki badań Fowlera, uściślenie terminów zawartych w pytaniu potrafi dać różnicę w rozkładach odpowiedzi wynoszącą od kilku do kilkudziesięciu punktów procentowych (1992: 226).

Druga zasadnicza wada pilotażu konwencjonalnego bierze się z tego, że ankieterzy realizują go na małych, okolicznościowych próbkach, a wywiad i obserwację zachowań respondenta prowadzą bez konkretnych wytycznych, czy – jak to ujęła K. Lutyńska – programu metodologicznego (1984: 59). Uzyskany w ten sposób materiał można określić jako przypadkowe obserwacje na przypadkowych respondentach. Dane zebrane bez przygotowania i ukierunkowanej perspektywy patrzenia są wyrywkowe i niesystematyczne, oparte na indywidualnych wrażeniach i nastawieniach. Doświadczenia z *debriefing interviewers* wskazują, że zgodność relacji ankieterów jest zwykle mała; problemy, które ujawniły się w trakcie pojedynczych „dramatycznych” wywiadów, są na ogół wyolbrzymiane (Oksenberg i in. 1991: 350). W efekcie przydatność konwencjonalnego pilotażu do diagnozowania i rozwiązywania problemów związanych z pytaniami jest niska – badacz nie uzyskuje godnych zaufania informacji, mówiących, które pytania są, a które nie są problematyczne, jak temu zaradzić, czy proponowane poprawki rzeczywiście uczynią pytanie bardziej miarodajnym.

Ostatecznie o pilotażu konwencjonalnym nie można powiedzieć, że jest techniką badawczą, czyli wzorem czynności ujętych w formie dyrektyw mówiących jak realizować badanie. Przedstawione powyżej „procedury” są jedynie rekonstrukcją praktyki badawczej, nieposiadającej metodologicznego uzasadnienia.

3. Pilotaż pogłębiony

Jeżeli chodzi o polskie piśmiennictwo naukowe, opracowana przez Krystynę Lutyńską koncepcja pilotażu pogłębionego jest zasadniczym wkładem w rozwój metod testowania narzędzi w badaniach kwestionariuszowych. Koncepcję tę wdrażano w kierowanej przez profesor Lutyńską Sekcji Badań Pilotażowych w Zakładzie Metodologii Badań Socjologicznych IFiS PAN w Łodzi (Gostkowski

1975: 8), realizując tam zarówno pilotaże, jak i prowadząc konsultacje metodologiczne narzędzia w powiązaniu z odpowiadającą mu koncepcją badawczą (Lutyńska 1984: 9, 56–58).

Przed skrótowną prezentacją tej propozycji należy dodać, że jej zapleczem teoretycznym, wypracowanym w „łódzkiej szkole metodologicznej”, jest tzw. rozszerzona informacyjna koncepcja pytania (Lutyński 1979). W pilotażu pogłębionym wyraźne są też wpływy koncepcji weryfikacji wewnętrznej i normatywnego modelu procesu komunikowania się w wywiadzie kwestionariuszowym (Lutyński 1975).

Zgodnie z założonymi celami pilotażu pogłębionego, wypróbowanie narzędzia badawczego (kwestionariusza, instrukcji dla ankieterów, materiałów pomocniczych) ma prowadzić do ustalenia: 1) Czy pytania kwestionariusza przynoszą odpowiedzi istotne, tzn. przynoszą poszukiwane przez badacza jednostkowe informacje?³; 2) Czy można mieć zaufanie do tego, że uzyskane informacje są prawdziwe, to znaczy czy odpowiedź jest nie tylko istotna, ale także trafna? (Lutyńska 1984:59). Niespełnienie w zadowalającym stopniu oczekiwań dotyczących zdolności pytania do dostarczania odpowiedzi istotnych i trafnych traktuje się jako wskazanie do jego przeredagowania lub usunięcia. Przy okazji można wspomnieć, że pilotaż realizowany według omawianej koncepcji pozwala także na zidentyfikowanie pytań niepoprawnych⁴. Pytań

³ Ustalenie istotności odpowiedzi na pytanie kwestionariusza (np. *Czy, P. zdaniem, wprowadzenie w Polsce euro będzie korzystne czy niekorzystne dla polskiej gospodarki?*) wymaga znajomości jednostkowego pytania badacza (*Czy, wg respondenta, wprowadzenie w Polsce euro będzie korzystne czy niekorzystne dla polskiej gospodarki?*). Wymaganie to podyktowane jest tym, że nie zawsze, tak jak w powyższym przykładzie, pytanie kwestionariuszowe jest pytaniem bezpośrednim, będącym tylko gramatycznym przekształceniem pytania badacza. Potrzeba operowania pojęciem istotności odpowiedzi wynika ze stosowania pośrednich pytań kwestionariuszowych, czyli takich, w których badacz bardziej lub mniej kamufluje przed respondentem intencję pytania, sugerując poszukiwanie innych informacji niż w rzeczywistości. Powyższe pytanie do respondenta mogłoby być uznane za pośrednie, jeśli towarzyszyłoby mu pytanie badacza *Czy respondent posiada opinię na temat skutków wprowadzenia euro dla polskiej gospodarki?* Ten prosty przykład pokazuje akurat problem odpowiedzi „trudno powiedzieć”, którą w pierwszym przypadku można uznać za nieistotną, a w drugim za istotną.

W wielkim uproszczeniu, o nieistotności odpowiedzi z r e g u ł y będą świadczyć takie zachowania respondenta, jak: milczenie, odpowiedzi „nie rozumiem”, odmowy odpowiedzi, wypowiedzi bez związku z poruszonym tematem, czyli niezgodne z podstawą pytania, a w przypadku pytań zamkniętych – niemieszczące się w przewidzianym zbiorze odpowiedzi do wyboru. Kategoria odpowiedzi nieistotnych potencjalnie obejmuje też wszystkie inne przypadki, w których respondent nie wykonuje zadania określonego w pytaniu, np. dokonuje mniejszej albo większej liczby wyborów odpowiedzi niż dopuścił do badacz.

⁴ Pytanie określa się jako niepoprawne, jeżeli spełnia ono co najmniej jeden z poniższych warunków: 1) poszukiwana informacja (pytanie badacza) nie została sformułowana; 2) w pytaniu kwestionariuszowym lub w pytaniu badacza jest błąd logiczny; 3) poszukiwana informacja nie dotyczy respondenta; 4) pytanie nie zostało zadane zgodnie z instrukcją (Lutyńska 1975: 58–60, Lutyńska 1983: 242).

tych nie kwalifikuje się jednak ze względu na istotność i trafność dostarczanych przez nie odpowiedzi (Lutyńska 1975: 58), co oczywiście nie wyklucza możliwości ich poprawy.

Zaleca się, aby realizację powyższych celów rozpocząć od konsultacji metodologicznych, które mogłyby stanowić pierwszy, choć niekonieczny etap pilotażu pogłębionego (Lutyńska 1984: 58). Pilotaż właściwy jest natomiast pomyślany jako badanie terenowe, w którym materiały zbierane są za pomocą „pogłębionych” wywiadów.

Koncepcja zakłada, że wywiady te będą przeprowadzane na małej – liczącej 30–100 respondentów – nielosowej próbie. Dobór jednostek jest celowy, według kryteriów wskazanych przez badacza i dopasowanych do specyfiki każdego projektu. Podkreślano jednak, że próba powinna być reprezentatywna typologicznie, przedstawiciele każdej kategorii społecznej znajdują w niej swój udział. Zwracano szczególną uwagę na kryterium wykształcenia, zalecając, by w próbie znalazły się osoby ze skrajnymi jego poziomami (Lutyńska 1984: 60–61).

Otwarta formuła pilotażu pogłębionego przewiduje możliwość, zgodnie z którą różne badania pilotażowe będą realizować nieco inaczej cele główne bądź nieco inne cele szczegółowe. Z tego między innymi powodu wymagano przygotowania osobnej instrukcji metodologicznej, dookreślającej zamierzenia i używane procedury badawcze (zob. Daniłowicz i Sztabiński 1974).

Określenie wywiadu pilotażowego mianem „pogłębionego” bierze się z tego, że sposób zbierania materiałów łączy w sobie cechy trzech technik: standaryzowanego wywiadu kwestionariuszowego, obserwacji oraz wywiadu o wywiadzie (por. Lutyńska 1984: 174). Wykorzystanie elementów tej ostatniej techniki polega na wywołaniu u respondenta dodatkowej wypowiedzi związanej z odpowiedzią udzieloną na pytanie kwestionariusza. Wypowiedź tę wywołuje się za pomocą tzw. pytania pilotażowego, będącego rodzajem pytania sondującego, czyli żądającego pogłębienia wcześniej udzielonych informacji.

Ogólne zasady prowadzenia wywiadu pilotażowego, mogące podlegać pewnym modyfikacjom określanym przez instrukcję metodologiczną, można, w odniesieniu do jednego testowanego pytania kwestionariuszowego, przedstawić następująco: na początku ankiet prowadzi wywiad według reguł typowych dla zasadniczego badania⁵ (Lutyńska 1984: 184–185), a w przypadku podejrzenia udzielenia odpowiedzi nietrafnej, zadaje pytanie (a) pilotażowe (Lutyńska 1984: 186). Od ankietera oczekuje się ponadto, że będzie on prowadził ciągłą obserwację zachowań respondenta, jego stanu psychofizycznego, obserwację sytuacji społecznej wywiadu i w końcu obserwację własnych zachowań i reakcji (w celu kontrolowania wpływu ankierskiego) (Lutyńska 1983: 230). Jako

⁵ Tzn. ankiet odczytuje pytanie kwestionariuszowe dokładnie w takim brzmieniu, jakie nadał mu badacz, a w przypadku otrzymania odpowiedzi nieistotnej, powtarza pytanie. Jeżeli mimo tego nie uzyskał odpowiedzi istotnej, przechodzi do kolejnego pytania w kwestionariuszu.

sposób utrwalenia materiałów oryginalnie stosowano zapis na piśmie (specjalnie przygotowane kwestionariusze wywiadu z jedną stroną poświęconą na notatki ankieterów; ankiety dla ankietera). Materiał ten, w przeważającej części jakościowy, ma dokładnie odtwarzać przebieg wywiadu w odniesieniu do każdego pytania kwestionariusza i uwzględniać przy tym możliwie wszystkie czynniki, które mogły wpłynąć na odpowiedź respondenta (Lutyńska 1984: 187–188). Z uwagi na te ponadprzeciętne wymagania określono, że cechy, jakimi powinien się odznaczać dobry ankieter-pilot, to: rzetelność, uczciwość, wysoki poziom inteligencji, a także zainteresowanie respondentem jako człowiekiem i tym dlaczego odpowiada w ten, a nie inny sposób (Lutyńska 1984: 197–198).

Rozwijając kwestię pytań pilotażowych, należy wspomnieć, że koncepcja zakładała dwa możliwe ich rodzaje. Pierwszy rodzaj to pytania znajdujące się w instrukcji metodologicznej, które ankieter powinien zadać w takiej formie i miejscu wywiadu, jak przewidział to badacz. Drugi rodzaj to pytania własne ankietera, formułowane przez niego w dowolny sposób. Ustalono, że pytanie takie ankieter może i powinien zadać zawsze wtedy, gdy uważa za słuszne sprawdzić trafność odpowiedzi respondenta. W ogólnej koncepcji pilotażu pozostawiono otwartą, do doprecyzowania w instrukcji metodologicznej, kwestię, kiedy zadawane są pytania pilotażowe – potencjalnie mogło to nastąpić przed, w czasie trwania lub po wywiadzie (Lutyńska 1984: 1985–1986). Analiza materiałów poglądowych (Lutyńska 1984: 200–205) sugerowałaby, że w praktyce pytania te zadawano w trakcie wywiadu.

Warto rozważyć teraz problem, w jakim zakresie poszczególne typy zbieranych informacji przyczyniać się miały do realizacji głównych celów pilotażu. I tak, ustalenie tego, czy pytanie przynosi odpowiedzi istotne, jest zadaniem stonkowo najprostszym i dającym rozstrzygnięcie niebudzące wątpliwości. Wystarczy do tego informacja o odpowiedzi respondenta na pytanie kwestionariusza (dająca się uzyskać w samym wywiadzie kwestionariuszowym, bez obserwacji czy dodatkowych pytań sondujących) oraz znajomość pytania badacza. Można tu zaznaczyć, że problem nieistotności odpowiedzi, a także problem przyczyn tej nieistotności został w ramach tej koncepcji uznany za mniej ważny niż problem trafności. Systematyczny sposób ustalenia przyczyn nieistotności odpowiedzi wymagałby zarówno danych pochodzących z obserwacji, jak i z dodatkowych pytań pogłębiających (Lutyńska 1983: 244–245). „Programowo” mniejsze zainteresowanie dla tego problemu może tłumaczyć, zawartą w ogólnych zasadach prowadzenia wywiadu pilotażowego, dyrektywę, aby ankieter zadawał własne pytanie pilotażowe jedynie w przypadku pojawienia się zastrzeżeń dotyczących trafności, a nie istotności odpowiedzi.

Natomiast sprawdzenie trafności odpowiedzi przedstawia się w sposób bardziej skomplikowany. Ocena odpowiedzi pod kątem tego, czy przynosi informację prawdziwą, pozostaje jedynie przypuszczeniem ankietera (stąd mówi się o odpowiedziach, w stosunku do których wysunięto zastrzeżenia co do ich trafności).

Podobnie wskazane przyczyny nietrafności można co najwyżej określić jako prawdopodobne. Taki status tych ustaleń wynika z tego, że opierają się one jedynie na obserwacji zachowań werbalnych i niewerbalnych respondenta oraz otoczenia sytuacji wywiadu (nie stosuje się metod weryfikacyjnych w oparciu o zewnętrzne źródła informacji). Wykorzystuje się przy tym założenie mówiące, że na podstawie zachowania się respondenta i jego wypowiedzi (w tym wywołanych pytaniem pilotażowym) można, częściowo przynajmniej, wnioskować o procesach psychicznych zachodzących u niego pod wpływem pytania kwestionariuszowego (Lutyńska 1984: 59, Lutyńska 1983: 245).

W tym miejscu warto podkreślić, wydawałoby się oczywistą, kwestię, że wszystkie zabiegi badawcze w pilotażu wykraczające poza czynności typowe dla badania zasadniczego, a więc zadawanie pytań pilotażowych i obserwacja, mają na celu nie tyle uzyskanie odpowiedzi trafnej, ile poznanie źródeł nietrafności. Jeżeli bowiem ostatecznym celem pilotażu jest przygotowanie nowej wersji kwestionariusza, która będzie dawała bardziej miarodajne rezultaty, to informacje o rozkładach odpowiedzi na pytanie kwestionariuszowe w badaniu pilotażowym będą do tego mniej przydatne niż ustalenia mówiące, jaką część nietrafności odpowiedzi można przypisać problemom z rozumieniem, wywołaniu emocji (takich jak np. strach, wstyd, mogących wskazywać na drażliwość pytania), a jaka część wynika z czynników sytuacyjnych (np. stanu psychofizycznego respondenta czy obecności osób trzecich), stanowiących słabszą przesłankę do przeredagowania lub usunięcia pytania.

Powyższa uwaga zawiera także pewne informacje odnośnie do opracowania materiałów. Koncepcja dopuszcza różne sposoby w tym zakresie, niemniej w wersji stosowanej przez pomysłodawczynię opracowanie to ma w dużej mierze charakter ilościowy. Jednostką agregacji jest pytanie kwestionariuszowe, w odniesieniu do którego należy podać przynajmniej takie podstawowe informacje, jak: liczba wystąpień odpowiedzi nieistotnych; istotnych bez zastrzeżeń co do trafności; istotnych z zastrzeżeniami co do trafności, a w przypadku tych ostatnich – częstość wystąpienia poszczególnych przyczyn nietrafności. Dane te oczywiście warto bogato zilustrować przykładami, gdyż mogą one dać wskazówki co do sposobu korekty pytań (Lutyńska 1984: 61, Lutyńska 1975: 56–64). Podejmując decyzje w tej kwestii, badacz powinien również uwzględnić rodzaj i liczebność próby, na jakiej zrealizowano badanie pilotażowe.

W podsumowaniu tej propozycji warto podkreślić, że została ona opracowana w sposób kompleksowy – zawiera ustalenia dotyczące technik doboru próby, zbierania materiałów oraz ich opracowania – z tego względu zasługuje na miano metody, czyli zespołu kompatybilnych technik. Biorąc to pod uwagę, określenie „koncepcja”, używane przez samą Autorkę, a dalej przeze mnie w artykule, wydaje się być zbyt słabe. Być może, przy szerokim rozumieniu słowa „koncepcja”, uzasadnieniem dla jego stosowania jest inna cecha pilotażu pogłębionego, jaką jest duża swoboda pozostawiona badaczom w zakresie określania celów

szczegółowych i sposobów ich osiągania – nie osłabia to jednak statusu pilotażu pogłębionego jako metody. Przykładowo, pilotaż pogłębiony nie narzuca reguł odnośnie do pytań pilotażowych badacza, nie wyklucza wykorzystania metod weryfikacji zewnętrznej (Lutyńska 1975a), stosowania pytań pustych (Daniłowicz 1975: 309) czy prowadzenia badań metodologicznych. W końcu też, co warto przypomnieć, F. Sztabińskiemu udało się zaadaptować „koncepcję” pilotażu pogłębionego do pilotażu ankiety indywidualnej i opracować na jego potrzeby nową technikę gromadzenia materiałów – ankietę pilotażową⁶ (1982).

4. Wywiady kognitywne

Popularność wywiadu kognitywnego (*cognitive interviewing*) rozpoczęła się w latach 80. XX w. w Stanach Zjednoczonych, gdzie stał się on dominującą techniką badań pilotażowych nie tylko w akademickich, ale także komercyjnych i publicznych ośrodkach badań sondażowych, podobnie jak w wielu instytutach statystyki publicznej (Beatty, Willis 2007: 288).

Technika ta podlegała rozwojowi i współcześnie ma wiele, czasami znacznie różniących się od siebie wariantów, które kryją się pod jedną nazwą „wywiad kognitywny”. Jest kwestią dyskusyjną, czy można traktować istniejące odgałęzienia jako tworzące grupę samodzielne techniki. W praktyce badawczej bywa bowiem, że bez większej refleksji są stosowane łącznie albo wymiennie. Poza tym dopiero niedawno podjęto próbę zmierzającą do ich opisu i systematyzacji w oparciu o zgromadzony dorobek badawczy (opracowanie Beatty’ego i Willisa 2007).

Cechą wspólną dla wszystkich typów wywiadów kognitywnych jest wywołanie u respondenta dodatkowej wypowiedzi związanej z odpowiedzią udzieloną na pytanie kwestionariusza. Wypowiedź tę traktuje się jako źródło informacji o poznawczych procesach zachodzących pod wpływem pytania kwestionariuszowego. Informacje te stanowią podstawę oceny pytania (zob. Beatty, Willis 2007: 288). Tym samym można powiedzieć, że ocena pytania dokonywana jest przede wszystkim przy uwzględnieniu perspektywy respondenta.

Potrzeba wywołania dodatkowych źródeł informacji wynika z tego, że w trakcie realizacji standaryzowanego wywiadu kwestionariuszowego procesy poznawcze pozostają w zasadzie nieuchwytne, dostępna jest jedynie informacja o ich produkcie końcowym. Wgląd w same procesy mają umożliwić właśnie wywiady

⁶ Procedura ankiety pilotażowej, podobnie jak wywiadu pilotażowego, zakłada wykorzystanie trzech źródeł informacji. Przebiega ona w następujący sposób: Badany wypełnia kwestionariusz w obecności ankietera; Ankieter w tym czasie obserwuje i notuje jego reakcje (werbalne i niewerbalne); Respondentowi uświadamiany jest pilotażowy cel badania – ma zaznaczyć pytania niejasne, z różnych względów trudne, jest także informowany, że po wypełnieniu kwestionariusza ankieter odbędzie z nim rozmowę (wywiad o ankiecie).

kognitywne, dzięki którym obserwacji będą mogły podlegać kolejne operacje poznawcze: 1) rozumienie pytania (jego intencji, terminów w nim zawartych); 2) wydobywanie z pamięci potrzebnych informacji (rodzaj informacji, strategie przypominania); 3) podejmowanie decyzji, formułowanie sądu (motywacja i włożony wysiłek, chęć ujawnienia prawdziwego sądu); 4) formatowanie odpowiedzi (dopasowywanie sądu do zaproponowanych przez badacza kategorii). Można tu dodać, że procesy poznawcze mogą nieco inaczej przebiegać w zależności od typu pytania kwestionariuszowego (zob. Willis 1999: 2).

Jak wspomniano, istnieją różne sposoby stawiające sobie za cel odsłonięcie wymienionych procesów poznawczych. Analizując wszystkie funkcjonujące praktyki, Beatty i Willis (2007) proponują wyróżnić dwa główne podejścia – paradygmaty, jak je, nieco przesadnie, nazywają autorzy. Pierwsze nosi nazwę technika myślenia na głos (*thinking-aloud*). Drugie podejście – *probing* – można określić mianem techniki pytań pogłębiających lub krócej – sondowania.

Technika myślenia na głos, w swej czystej postaci, polega na tym, że przed respondentem stawiane jest zadanie, aby na bieżąco werbalizował swoje myśli, w trakcie gdy odpowiada na pytanie kwestionariusza. Relacja ta jest najczęściej nagrywana (ewentualnie notowana). Rola ankietera jest mało aktywna – poza przedstawieniem wspomnianego zadania i odczytywaniem pytań z kwestionariusza, ma sprowadzać się do podtrzymywania u respondenta stanu, w którym werbalizuje on własne myśli i nie odbiega przy tym od tematu, nie dokonuje interpretacji oraz ocen. Zasada mówi, że ingerencja ankietera w wypowiedź respondenta powinna być możliwie jak najmniejsza (por. Beatty, Willis 2007: 289; por. van Someren i in. 1994: 30).

W tym miejscu warto zastanowić się nad źródłem pomysłu – powodem, dla którego zaczęto wykorzystywać ten rodzaju materiału. Można rozpocząć od tego, że technika *thinking-aloud*⁷ jest pierwotnym sposobem realizacji wywiadów kognitywnych, a jej początki przypadają na czas rozpoczęcia współpracy metodologów sondażowych z psychologami poznawczymi zajmującymi się problematyką rozwiązywania zadań przez człowieka. Bezpośrednich inspiracji dostarczyli w tym względzie Ericsson i Simons, którzy pokazali myślenie na głos jako wartościowe źródło informacji o zachodzących procesach poznawczych. Opierając się na rozległym dorobku empirycznym psychologii poznawczej, poddali testom własne hipotezy i znaleźli uzasadnienie m.in. dla twierdzenia mówiącego, że myślenie na głos, jako czynność dodatkowa, nie zmienia struktury ani toku procesów poznawczych związanych z realizacją podstawowego zadania – może je co najwyżej spowalniać. Jednocześnie werbalizacji podlegać może tylko zawartość pamięci krótkotrwałej, jako bezpośrednio dostępnej i „aktywnej” (Ericsson, Simons 1980: 223–226; van Someren i in. 1994: 26–33). Pożądany (z punktu widzenia użyteczności metody) brak wpływu werbalizacji

⁷ Oryginalnie nosiła nazwę *protocol analysis*.

na procesy poznawcze ma zachodzić głównie wtedy, jeśli podstawowe zadanie ma charakter werbalny. Przypadek wywiadu kwestionariuszowego generalnie spełnia ten warunek (pytanie kwestionariuszowe jako werbalny bodziec i odpowiedź jako werbalna reakcja), o ile nie wymaga się od respondenta dostarczenia odpowiedzi na podstawie analizy rysunku, diagramu itp. Jednak można wysunąć wątpliwość, czy taka interakcja nie zachodzi także w przypadku formowania odpowiedzi na pytania o opinie, jako procesu obejmującego emocje i preferencje (Conrad i in. 2000).

Zwolennicy tej techniki podkreślali, że myślenie na głos jako czynność równoległa do procesów poznawczych, w odróżnieniu od wszystkich relacji *ex post*, jest mniej podatne na różne formy zniekształceń wynikających z racjonalizacji własnych zachowań i przyjmowania strategii mających je uzasadniać w oczach odbiorcy (Loftus 1984: 61). Podobnie złagodzony jest problem artefaktualności danych, występujący na skutek oddziaływania ankietera – w przypadku *thinking aloud* rola ankietera jest stosunkowo bierna, a jego zachowanie w dużym stopniu zestandaryzowane. W efekcie, co przekłada się na kolejną zaletę, prowadzenie tego rodzaju wywiadów nie wymaga od ankietera posiadania jakichś specjalistycznych kwalifikacji (Beatty, Willis 2007: 292).

Do listy ograniczeń można dołączyć argument mówiący, że nie wszyscy respondenci, jak zauważono, radzą sobie równie dobrze z zadaniem myślenia na głos – u niektórych (mniejszości) czynność ta przebiega dość nieporadnie i dostarczają niepełnowartościowych materiałów. Jak się przypuszcza, owa nieporadność nie jest odzwierciedleniem („nieporadnego”) przebiegu procesów poznawczych, ale wynika z samych zdolności werbalizacyjnych. Nie odkryto istotnych zależności między tą umiejętnością a innymi cechami respondentów. Wyjątek stanowi wiek – u małych dzieci często pojawia się trudność z werbalizacją bieżących myśli (van Someren i in. 1994: 35–36). Jeżeli przypuszczenie jest prawdziwe, badacze mogliby przeprowadzać wywiady jedynie z osobami o dobrej zdolności werbalizacji, nie martwiąc się groźbą skrzywienia rezultatów na skutek selektywnego doboru jednostek.

Przy okazji warto nadmienić, że praktyką amerykańskich badaczy jest prowadzenie wywiadów kognitywnych w warunkach laboratoryjnych, a nie terenowych, np. w domach respondentów (Beatty, Willis 2007: 288). Niestety nie udało mi się odnaleźć w literaturze merytorycznego uzasadnienia dla tego rozwiązania. Nie jest więc pewne, czy chodzi tylko o techniczne udogodnienia, czy może przyjęto założenie, że myślenie na głos na prośbę obcej osoby, a więc czynność niestandardowa i dość sztuczna, będzie przebiegać efektywniej i przy lepszym nastawieniu respondenta, jeśli zostanie on przeniesiony na „teren badacza”. Bliżej na ten temat wypowiedział się jedynie Willis, którego zdaniem do przeprowadzenia wywiadu kognitywnego nie potrzeba specjalnego „laboratorium”, a jedynie cichego pomieszczenia, czy to na terenie badacza czy respondenta (Willis 1999: 26).

Wśród badaczy odnoszących się z rezerwą do techniki myślenia na głos pojawia się też argument, że relacje respondentów, z racji pozostawionej im swobody wypowiedzi, dostarczają mniej informacji i nie są bezpośrednio zogniskowane na wszystkich interesujących badaczy problemach – efekt ten można dopiero osiągnąć poprzez ukierunkowane sondowanie. Przykładowo uważa się, że materiały mają mniejszą użyteczność w identyfikowaniu przypadków niezrozumienia terminów w pytaniu. Natomiast największe zalety techniki *thinking aloud* zdają się wynikać z jej możliwości badania procesów przypominania, stąd jest ona godnym polecenia sposobem testowania pytań o fakty i doświadczenia z życia respondenta (por. Beatty, Willis 2007: 294). Poniżej są przedstawione dwa przykłady zastosowań w tym zakresie.

E. Loftus prowadziła badania na małych próbkach, w odniesieniu do pytań żądających od respondenta przypomnienia sobie pewnych faktów z jego życia („Biorąc pod uwagę ostatnie 12 miesięcy, jak często... /czy...?”). Ustaliła, po pierwsze, że u większości respondentów proces przypominania przebiegał od przeszłości do teraźniejszości (co niezgodne jest z oczekiwaniem, że wydarzenia nieodległe w czasie będą bardziej dostępne w pamięci). Po drugie, zwróciła uwagę na to, że w trakcie „głośnego myślenia”, któremu towarzyszy nieustanny proces przypominania, u respondentów obserwuje się zmiany początkowych deklaracji. Podobnie w pytaniach szczegółowych, drugich w kolejności po ogólnych, respondenci przypominają sobie więcej faktów, w efekcie czego odpowiedź na pierwsze pytanie należy uznać za nietrafną. W oparciu o te wstępne obserwacje (niekiedy wręcz banalne i zgodne z naszym potocznym doświadczeniem), Loftus postawiła pytania odnoszące się do tego, czy, ewentualnie jak, jesteśmy w stanie ustalenia te wykorzystać w praktyce do konstruowania pytań w „zwykłym” wywiadzie kwestionariuszowym, tak aby uzyskiwać bardziej miarodajne rezultaty. Może w pytaniu warto umieszczać „wskazówki” wspomagające proces odtwarzania z pamięci (np. „Biorąc pod uwagę okres od stycznia zeszłego roku do stycznia tego roku...?”)? Może należałoby zostawić respondentowi dłuższy czas do namysłu, np. 1 minutę, informując go o tym w pytaniu? Może należałoby zrezygnować ze stawiania pytań ogólnych przed szczegółowymi? (zob. Loftus 1984: 61–63).

Inny przykład korzyści z zastosowania techniki myślenia na głos pochodzi z badań prowadzonych przez DeMaio, Ciochetto i Davis (1993: 1021–1024) przy okazji realizacji sondażu na temat zachowań żywieniowych, w którym zbierano od respondentów drobiazgowo informacje na temat wszystkich przyjętych przez nich poprzedniego dnia pokarmów i napojów. Zaobserwowano, że ułożenie pytań w sposób odtwarzający porządek godzinowy (kiedy → w ramach jakiego posiłku → co → pozostałe żądania opisu) jest mało efektywne, gdyż nie odpowiada dominującym strategiom przypominania. Po przeprowadzonych wywiadach kognitywnych i przetestowaniu wprowadzonych w następstwie poprawek,

badacze ułożyli pytania w sekwencji rozpoczynającej się od „co”. Zauważyli bowiem, że w procesie przypominania respondenci nie posługują się wskazówkami na temat godziny czy typu posiłku (śniadanie czy obiad itp.), ale zwykle „myślą” nazwą jedzenia lub konkretnymi zdarzeniami w ich wczorajszym dniu, którym towarzyszył posiłek (zachowując przy tym chronologię zdarzeń). Inną przyjętą poprawką było zadanie kilku różnych pytań „co” (odnoszących się do tej samej poszukiwanej informacji), tak aby wydłużyć i wspomóc proces przypominania.

Zupełnie innym sposobem realizacji wywiadów kognitywnych jest **technika sondowania** (*probing, intensive interview*). Źródłem informacji o procesach poznawczych są tu wypowiedzi respondenta na temat wcześniej udzielonej odpowiedzi na pytanie kwestionariusza, uzyskane za pomocą dodatkowych pytań sondujących (*cognitive probes*). Jak widać, technika ta zakłada dwa etapy. W pierwszym symuluje się typowe warunki wywiadu kwestionariuszowego, w których ankietę odczytuje testowane pytanie kwestionariusza, a respondent udziela na nie odpowiedzi. W drugim etapie ankietę przystępuje do zadawania pytań pogłębiających.

W przeciwieństwie do *thinking-aloud*, relacja respondenta ma charakter retrospektywny, a nie bieżący. W nowszej literaturze podkreśla się, że retrospekcja powinna następować bezpośrednio po udzieleniu odpowiedzi na konkretne pytanie kwestionariusza (a nie całej ich baterii), tak aby respondent miał jeszcze dostęp do informacji przechowywanych w pamięci krótkotrwałej, tych samych, z których korzystał formułując odpowiedź na pytanie kwestionariusza. To postępowanie ma zapewnić trafność dokonywanej rekonstrukcji procesów poznawczych (zob. Presser i in. 2004: 112). Podkreślając inne różnice wobec techniki myślenia na głos, trzeba również zauważyć, że w przypadku *probing* wypowiedź nie jest swobodnym monologowaniem respondenta, „potokiem myśli”, ale efektem rozmowy kierowanej przez ankietę. Jego rolę można więc określić jako generalnie aktywną, szczególnie jeśli ma pozostawioną dużą swobodę w zakresie zadawania pytań, co wymaga ciągłej i uważnej analizy wypowiedzi respondenta (Beatty, Willis 2007: 289). Przebieg całego procesu interwencji zwykle utrwalą się w formie nagrania.

Ewolucja wywiadu kognitywnego, polegająca na stopniowym odchodzeniu od myślenia na głos na rzecz wywiadu ukierunkowanego za pomocą pytań sondujących, miała przede wszystkim przesłanki praktyczne – tę ostatnią technikę oceniano jako przynoszącą więcej zróżnicowanych informacji przydatnych do oceny kwestionariusza. Nie stało to zresztą w sprzeczności z refleksją teoretyczną – w świetle przekonań Ericssona i Simonsa takie rozwiązanie ma swoje uzasadnienie, dając – przynajmniej częściowy – ciągły wgląd w procesy poznawcze (por. 1980: 226).

Przechodząc do dokładniejszej charakterystyki sposobu zbierania materiału, należy zaznaczyć, że pytania pogłębiające mogą być zarówno ogólne,

wywołujące podobny, pod pewnymi względami, typ narracji jak przy myśleniu na głos (np.: Co wtedy myślałeś? Jak dochodziłeś do tej odpowiedzi? Powiedz coś więcej na ten temat), jak i bardziej szczegółowe, odnoszące się do wybranych aspektów procesu poznawczego (np.: Co to słowo dla Ciebie znaczy? Czy mógłbyś własnymi słowami powiedzieć o co chodzi w tym pytaniu?; Jak pamiętasz...? Jak sobie przypominałeś...?; Na ile byłeś pewny, że ...?) (zob. Willis 1999: 6). Bliższy ogląd praktyk badawczych pokazuje, że treść zadawanych pytań sondujących nie jest w tej technice ustalona na drodze jakiejś obowiązującej definicji, a proces interrogacji nie musi przebiegać według jednolitego wzorca. Co się z tym bezpośrednio wiąże, zróżnicowane są style prowadzenia wywiadu, od proaktywnego – w którym o treści pytań decyduje badacz lub ankieter kierujący się wyznaczonym wcześniej konkretnym zapotrzebowaniem informacyjnym, do reaktywnego – w którym pytania są zadawane w oparciu o bieżącą obserwację zachowań respondenta i tylko w przypadku wystąpienia problemów z udzielaniem odpowiedzi na pytanie kwestionariusza. Podobne zróżnicowanie występuje w odniesieniu do formy pytań – jakkolwiek są one w zdecydowanej większości pytaniami otwartymi, żądającymi od respondenta dłuższej narracji, to ich stopień standaryzacji może być zarówno niski (brzmienie ustalane przez ankietera, dopasowywane każdorazowo do respondenta i przebiegu wywiadu), jak i wysoki (jednolite brzmienie ustalone *ex ante* przez badacza, narzucone do stosowania każdemu ankieterowi) (por. Beatty, Willis 2007: 300).

Sposób zadawania pytań pogłębiających może się zmieniać w trakcie wywiadu i być dopasowywany do poszczególnych pytań kwestionariusza. Praktyka badawcza wskazywałaby, że stosowanie stylu proaktywnego jest zasadne w przypadku, kiedy badacz podejrzewa, że pytanie może stwarzać problemy. Obejmuje to między innymi przewidywanie co do zajścia „milczących nieporozumień”, czyli sytuacji, w których respondenci nie zdradzają swoim zachowaniem żadnych problemów z dostarczaniem odpowiedzi, podczas gdy takowe występują (niezrozumienie pytania zgodnie z intencją badacza). Podkreśla się jednak, że wskazania do przyjęcia konkretnego sposobu sondowania powinny być dookreślone na podstawie dalszych badań metodologicznych, mających na celu ustalenie, jak często i jakie problemy dotyczące określonego pytania ujawniają się, przy jakich stylach prowadzenia wywiadu (Beatty, Willis 2007: 300–301).

Obydwie odmiany wywiadu kognitywnego – technika myślenia na głos i technika sondowania – uzupełniają się z grubsza pod względem zalet i wad. Jeżeli w danej dziedzinie jedna technika ma ograniczenia (czy zalety), druga ich nie posiada. W *probing* procesy poznawcze oraz ich werbalizacja są czynnościami następującymi po sobie, a więc relacja o procesach nie może wpływać na te procesy – w przypadku *thinking aloud* nie ma co do tego pewności. W technice myślenia na głos „odpowiedzialność” za niewielką użyteczność

materiału jest przerzucona na zdolności werbalizacyjne respondenta i częściowo metodę, a w przypadku techniki pytań sondujących powodzenie w dostarczeniu cennych informacji zależy od ankietera i jego umiejętności zadania właściwych pytań. Pozostawienie ankieterowi swobody w prowadzeniu wywiadu potencjalnie rodzi niebezpieczeństwo sytuacji, w której ustalenie „problemu z pytaniem” będzie tylko produktem interakcji ankietera z respondentem – zagrożenie to jest zminimalizowane przy technice *thinking aloud*. W końcu *probing* może diagnozować więcej rodzajów problemów z pytaniem niż mająca mniejsze, acz bardziej wyspecjalizowane zastosowanie, technika myślenia na głos.

Niezależnie od tego, czy techniką zbierania materiałów jest myślenie na głos czy sondowanie (w którymś ze swych wariantów), wywiady kognitywne przynoszą materiały jakościowe. Jak można się domyślać, nie istnieje tylko jeden sposób ich opracowania. Sytuacja ta jest konsekwencją zróżnicowanego, do pewnego stopnia, zapotrzebowania informacyjnego badaczy oraz nieco innych możliwości poszczególnych technik w odsłanianiu defektów pytania (por. Beatty, Willis 2007: 301). Można zasygnalizować, że próbę stworzenia schematu opracowania, ale głównie do badań o charakterze ilościowym, podjęli na przykład Conrad i Blair (1996), opierając się na modelu przebiegu procesów poznawczych (w skróconej wersji). Kierunek, w jakim powinny być skonceptualizowane techniki opracowania materiałów (jakościowy vs ilościowy), nie jest jednak przesądzony, a na taką decyzję wpływa także sposób doboru próby.

W praktyce dominują niewielkie próby celowe, składające się od kilku do kilkudziesięciu respondentów. Beatty i Willis postulują, aby tę kwestię ostatecznie uregulować i wprowadzić odpowiednie standardy. Według nich idealną praktyką byłoby prowadzenie badania w rundach składających się z 5–15 wywiadów i kontynuowanie tej procedury do czasu, w którym kolejne rundy przestaną ujawniać nowe problemy z pytaniem. Rozwiązanie to podyktowane jest obserwacją wskazującą, że do zdiagnozowania najważniejszych problemów stwarzanych przez pytanie nie zawsze wystarczy 50 wywiadów, a problemy te potrafią się ujawniać dopiero pod koniec realizacji pilotażu (2007: 295–296, 306).

Interpretacja wyników uzyskanych za pomocą wywiadów kognitywnych musi też uwzględniać fakt, że dane uzyskuje się na drodze procesu interwencji, który znacząco odbiega od warunków realizacji wywiadu w badaniu zasadniczym. Jeżeli badacze realizujący pilotaż decydują się na taki „brak realizmu”, to dlatego, że spodziewają się uzyskać w zamian dodatkowe informacje pozwalające ocenić narzędzie. Niestety zawsze ryzykują przy tym, że zidentyfikują problemy, których nie będzie w wywiadzie zasadniczym (*errors of commission*) i nie zidentyfikują tych, które w nim wystąpią (*errors of omission*) (Beatty, Willis 2007: 302).

5. Techniki dodatkowe w pilotażu konwencjonalnym oraz jego modyfikacje

Podczas kiedy zastosowanie wywiadów kognitywnych czy pilotażu pogłębianego zawsze pociąga za sobą organizację odrębnego typu badania, poniżej omówione techniki potencjalnie można stosować także w badaniach zasadniczych, gdyż ich użycie nie zaburza toku standaryzowanego wywiadu. Jeżeli mają być one wykorzystane z myślą o przetestowaniu narzędzia, to oczywiście musi to nastąpić w ramach badań wstępnych, jednakże badania takie wykorzystują zazwyczaj te same źródła informacji, które wywoływane są w badaniach zasadniczych, natomiast różnica polega jedynie na użytku, jaki robi się z tych źródeł. W badaniach wstępnych (pilotażowych) chodzi o większe ich wykorzystanie poprzez systematyczną i bardziej drobiazgową analizę procesu udzielania odpowiedzi przez respondenta. Jednocześnie każda z tych technik jest zdolna do identyfikowania ograniczonej klasy problemów i w efekcie są one traktowane jako techniki dodatkowe, doskonalące konwencjonalne podejście.

Jedną z nich jest **ilościowa obserwacja zachowań werbalnych** (*behavior coding*). Jako technika używana do celów pilotażowych została opracowana przez Oksenberga, Cannella i Kaltona (1991). Zakłada, że obserwacja ma odnosić się zarówno do zachowań ankietera, jak i respondenta, przy czym – jak wskazuje moja propozycja polskiego odpowiednika oryginalnej nazwy – w polu zainteresowania pozostają tylko reakcje werbalne.

Przebieg wywiadu może być obserwowany na bieżąco (w trakcie trwania) lub po utrwaleniu w formie audialnej, czy video-audialnej (ewentualnie z transkrybowaniem). *Behavior coding* jest pomyślana jako technika ilościowa, dlatego przebieg wywiadu w odniesieniu do każdego pytania kwestionariuszowego poddawany jest najpierw kategoryzacji.

Oryginalnie technika zakłada, że kodowanie jest przeprowadzane przez „zewnątrznych” obserwatorów (dokonujących obserwacji na bieżąco lub pracujących na nagraniach) – ankieterzy prowadzący wywiad nie występują w tej roli. Jakkolwiek jest to rozwiązanie z wielu względów zasadne, wymaga zgody na nagranie lub obecności dodatkowej osoby w trakcie wywiadu. Zastosowanie tego ostatniego wariantu nastręcza dodatkowe problemy w wywiadach prowadzonych w terenie (np. w domach respondentów), natomiast dzięki możliwości włączenia podsłuchu jest bez trudu osiągalne w badaniach CATI.

W praktyce stosowane przez badaczy schematy kodowania mogą mniej lub bardziej różnić się od siebie, z reguły jednak czerpią z „klucza” opracowanego przez pomysłodawców techniki. Zachowanie ankietera jest w nim przedmiotem zainteresowania jedynie w odniesieniu do zmiennej – sposób odczytania pytania kwestionariuszowego: A) dokładnie jak w kwestionariuszu; B) odczytano niedokładnie, jednak bez zmiany znaczenia treści pytania; C) odczytano niedokładnie,

zmieniając znaczenie treści pytania. Warto podkreślić, że w odniesieniu do odpowiedzi respondenta autorzy posługują się pojęciem istotności i trafności odpowiedzi (w znaczeniu nadanym przez Lutyńskiego i przedstawionym wyżej). Za pomocą schematu ma być rejestrowane ewentualne wystąpienie (wystarczy, że raz w procesie udzielania odpowiedzi na jedno pytanie kwestionariuszowe) następujących sytuacji: 1) udziela odpowiedzi zanim ankieter skończy czytać pytanie; 2) prosi o powtórzenie pytania lub jego wyjaśnienie lub wygłasza wypowiedź wskazującą na brak pewności co do rozumienia pytania; 3) udziela odpowiedzi istotnej; 4) udziela odpowiedzi istotnej, odnośnie do której można wysunąć zastrzeżenie pod kątem trafności; 5) udziela odpowiedzi nieistotnej; 6) udziela odpowiedzi typu „trudno powiedzieć”; 7) odmawia udzielenia odpowiedzi (Oksenberg i in. 1991: 352–353). Jak widać, schemat do kodowania zachowań respondenta uwzględnia całą sekwencję jego ewentualnych wypowiedzi – tłumaczy to możliwość współwystępowania pewnych zachowań (np. udzielił odpowiedzi nieistotnej, a potem istotnej), które byłyby sprzeczne, gdyby miały określać pojedynczą wypowiedź.

Na 10 przewidzianych kodów (A–C, 1–7), 8 to wskaźniki problematyczności pytania (ang. *indicates a potential problem with the question*) – poza tym zbiorem pozostają kategorie A oraz 3. Próbuje się określać, jaka wartość wskaźnika świadczy o wysokiej problematyczności pytania, ustalając ją z reguły arbitralnie na poziomie 15% respondentów (Oksenberg i in. 1991: 353, Fowler 1992: 221). Warto dodać, że decyzja co do akceptowalnego poziomu tych wskaźników powinna uwzględniać także rodzaj doboru próby, który w tej technice nie jest z góry narzucony (por. Oksenberg i in. 1991: 351).

Można też wspomnieć o propozycjach innych schematów kodowania. Stosunkowo najprostsza i najwygodniejsza do analiz byłaby propozycja van der Zoowena i Smita, którzy przewidują tylko jedną zmienną do opisu sekwencji zachowań towarzyszących jednemu pytaniu kwestionariuszowemu. Wzajemnie rozłączne wartości tej zmiennej to: 1) modelowa – ankieter poprawnie odczytał pytanie; respondent dokonał wyboru alternatyw zgodnie z zadaniem; ankieter poprawnie zakodował pytanie w kwestionariuszu; 2) problematyczna – sekwencja nie przebiegała modelowo, ale problem został rozwiązany, np. respondent poprosił o wyjaśnienie i dokonał wyboru alternatywy w pytaniu; 3) nieadekwatna – sekwencja nie przebiegała modelowo, problem nie został rozwiązany. Zgodnie z wynikami badań autorów tego schematu, identyfikował on problematyczne pytania na poziomie porównywalnym do innych stosowanych schematów (2004, cyt. za: Presser i in. 2004: 116).

Podobne rozwiązania (z tą różnicą, że rolę obserwatorów pełnili ankieterzy) były wykorzystywane przez polskich badaczy. W Europejskim Sondażu Społecznym 2004 część procedur pilotażowych przewidywała, że dla każdego pytania będą zbierane informacje na temat reakcji respondenta według następującego klucza: 1z) objawy znużenia lub zniecierpliwienia respondenta; 2t) trudność

pytania; 3n) niejasność użytych w pytaniu terminów lub zwrotów; 4k) problemy z kafeterią (skalą) w pytaniu; 5pp) powtórzenie pytania; 6d) pytanie drażliwe; 7i) pytanie interesujące; 8iu) inne uwagi; 9bu) bez uwag (Sztabiński, Sztabiński 2005: 62–63). Można na marginesie wspomnieć, że w polskich badaniach prekursorskim narzędziem służącym prowadzeniu ilościowej obserwacji przebiegu wywiadu (głównie zachowań ankierów) był tzw. system S, opracowany przez J. Lutyńskiego (Daniłowicz 1974).

Wśród zalet tej techniki wymienia się przede wszystkim jej obiektywność, której sprzyja odsunięcie ankietera od czynności kodowania. Ponadto – jak wskazują wyniki badań – zapisy koderów odznaczają się wysokim poziomem rzetelności (między zapisami dwóch niezależnych zespołów koderów jest duża zgodność) (Oksenberg i in. 1991: 353; Esposito i in. 1991: 47).

Generalnie technika nie generuje też dużych kosztów. Praca polegająca na kodowaniu nie jest specjalnie obciążająca ani czasochłonna – prosty schemat klasyfikacyjny powoduje, że do zakodowania materiału powinno wystarczać jednorazowe przesłuchanie taśmy bez jej zatrzymywania (w analogicznych warunkach muszą pracować „naoczni” obserwatorzy), a dodatkowe wykorzystanie programu komputerowego umożliwia natychmiastowe otrzymanie wyników, które następnie mogą się stać podstawą do dyskusji, wymiany dodatkowych spostrzeżeń i interpretacji w trakcie spotkania badacza z ankieterami (Oksenberg i in. 1991: 362).

Nie bez praktycznego znaczenia pozostaje też fakt, że *behavior coding* to technika, która w ogóle dostarcza konkretnego narzędzia, na dodatek prostego, dającego się zaaplikować bodaj w każdym badaniu pilotującym kwestionariusz wywiadu oraz generującego materiały dające się łatwo i szybko opracować.

To podejście ma jednak swoje wady. Najpoważniejsze ograniczenia polegają na tym, że technika nie jest zdolna identyfikować ani wszystkich problemów, ani źródeł wszystkich rozpoznanych problemów. W tym punkcie technika obserwacji zachowań werbalnych dzieli swe ograniczenia ze wszystkimi podejściami, które, mając na celu przetestowanie narzędzia, opierają się jedynie na źródłach informacji typowych dla badań zasadniczych. Szczególnie dotyczy to problemów z rozumieniem pytania, które aby być identyfikowalne muszą być w tym przypadku wyraźnie zasygnalizowane przez samego respondenta. Z kolei brak możliwości zadawania pytań sondujących utrudnia odsłonięcie sposobu, w jaki pojmowane jest pytanie. Słaby wgląd w wewnętrzne procesy zachodzące u respondenta to także cena, jaka wynika z przyjęcia zewnętrznej perspektywy obserwacji, stawianej na uchwycenie tego co łatwo mierzalne.

O *behavior coding* można inaczej powiedzieć, że bada przede wszystkim sprawność przebiegu wywiadu – wychwytuje te pytania, które zaburzają jego gładki tok i w tym sensie są problematyczne, czy to dla respondenta czy dla ankietera. Usunięcie tak rozumianych trudności nie gwarantuje jednak, że pytania przestały być źródłem błędów i przynoszą rezultaty, do których można mieć zaufanie.

Z tego powodu obserwacja zachowań werbalnych nie powinna być w badaniu pilotażowym jedyną stosowaną techniką, czego pełną świadomość mieli sami jej autorzy (Oksenberg i in. 1991).

W polu zainteresowania metodologów pozostają też pozatreściowe aspekty przebiegu wywiadu – zachowania niewerbalne i parajęzykowe. Jeden z kierunków intensywnie prowadzonych badań nastawiony jest na obserwację respondenta pod względem długości czasu, jaki zajmuje mu udzielenie odpowiedzi na pytanie kwestionariusza (*response latency*).

Potencjał tej informacji dostrzegła już K. Lutyńska i wykorzystała ją jako wskaźnik oceny trudności oraz drażliwości pytań (badanie odnosiło się konkretnie do pytań metryczkowych). Z racji braku możliwości zastosowania dokładnych urządzeń pomiarowych, odnotowywano jedynie, czy respondent odpowiadał szybko i bez namysłu czy po dłuższym zastanowieniu. Analizy prowadzono zgodnie z założeniem, że pytaniom trudnym (albo drażliwym) towarzyszy większy namysł (1978).

Obecnie prowadzone badania nad „opóźnieniem w dostarczeniu odpowiedzi” zakładają wykorzystanie takich urządzeń, które dają pomiar ilościowy z bardzo dużą dokładnością. Biorąc pod uwagę tylko te przedsięwzięcia, które mogą przynieść wkład w udoskonalenie praktyk pilotażowych, kierunki poszukiwań i tak są dość zróżnicowane. Przykładowo Bassili i Scott (1996) podjęli badania, z myślą że *response latency* może okazać się narzędziem zdolnym wskazywać pytania wadliwe. W tym celu porównali średni czas namysłu nad pytaniami wadliwymi (z wadami logicznymi oraz wadami ustalonymi na drodze empirycznej za pomocą *behavior coding*) i ich poprawnymi odpowiednikami. Niejednoznaczne wyniki ich badania pokazują, że pytania wadliwe nie zawsze wywołują dłuższy czas formowania odpowiedzi – „opóźnienie” może nastąpić także po ich korekcie (jak w tym przypadku przypuszczają badacze, po doprecyzowaniu znaczenia terminów w pytaniu doszło do – zajmującego czas – pełniejszego uruchomienia procesów pamięciowych).

Jeszcze innym tropem poszli Draisma i Dijkstra, którzy mogli porównać odpowiedzi kwestionariuszowe z danymi ze źródeł zewnętrznych – zaobserwowali, że odpowiedziom nieprawdziwym towarzyszą oznaki niepewności respondenta oraz wydłużanie się czasu potrzebnego do odpowiedzi. Ustalenie to mogłoby być pośrednio wykorzystywane w badaniach pilotażowych do oceny pytań, a konkretnie ich zdolności do dostarczania wiarygodnych odpowiedzi. Niestety, te obserwacje nie znalazły potwierdzenia w innych badaniach, zatem za wcześnie, aby traktować *response latency* jako uniwersalny wskaźnik nietrafności odpowiedzi (2004, cyt za: Presser i in. 2004: 116–117). Jak sugerowały to chociażby badania Bassiliego i Scotta, respondent może pożytkować czas na dokładniejsze „przemyślenie”, sprzyjające dostarczeniu miarodajnej odpowiedzi.

Zanim *response latency* stanie się w pełni dopracowaną techniką, rutynowo stosowaną w badaniach pilotażowych, musi zostać lepiej zbadana – jak bowiem

pokazano, nie jest w pełni jasne, czego wskaźnikiem jest „opóźnienie odpowiedzi”. Do doprecyzowania pozostają też kwestie definicyjne i operacyjne – kiedy rozpoczyna się „czas latencji” (zaraz po odczytaniu pytania przez ankietera, czy też jeszcze w trakcie odczytywania, skoro już wtedy procesy poznawcze mogą się uruchomić); kiedy kończy (w momencie, kiedy respondent rozpoczyna, czy kiedy kończy odpowiedź, którą zapisuje ankieter); jakie są jego składniki (tylko milczenie, czy także inne wypowiedzi pomiędzy pytaniem a ostateczną odpowiedzią) (zob. Bassili, Scott 1996: 392, 394–395; zob. Schaeffer i in. 2008: 4344).

Jeszcze inną techniką są **winiety**. Ich wykorzystanie w badaniach społecznych jest bardzo szerokie. Przedstawiając skrótowo tę technikę i jej najprostsze zastosowanie w pilotażu, można powiedzieć, że służy ona do badania (i rozwiązywania) problemów związanych z rozumieniem przez respondentów terminów i zwrotów zawartych w pytaniach kwestionariuszowych. Jak wiadomo, nawet będące w powszechnym użyciu pojęcia czy wyrażenia, takie jak „praca”, „szukać pracy”, „z wyłączeniem prac domowych”, nie są wystarczająco jednoznaczne i mają swoje różne konotacje. Jednocześnie przekazanie respondentom znaczenia, o które chodzi badaczowi, nie jest sprawą łatwą także z tego powodu, że pytanie kwestionariuszowe, za pośrednictwem którego to się dzieje, powinno być względnie krótkie i „wystandaryzowane” do poziomu przeciętnego respondenta.

Problemy te złagodzić można poprzez umieszczenie w badaniu pilotażowym zarówno testowanego pytania, jak i ułożonych do niego winiety. Winiety, podobnie jak „historyjki”, to proste i konkretne opisy hipotetycznych sytuacji lub osób, będące – mówiąc językiem bardziej sformalizowanym – różnymi operacyjnymi definicjami terminu teoretycznego użytego w testowanym pytaniu. Po każdej winiecie, do respondenta kierowane jest pytanie służące ustaleniu, czy opisana sytuacja odpowiada znaczeniu, jakie respondent nadaje danemu słowu (np. Czy czynności te należą, P. zdaniem, do prac domowych?). Dodać trzeba, że testowane pytanie jest zadawane w podstawowej części wywiadu, a winiety po jego zakończeniu. Jakkolwiek więc technika winiety zakłada wywołanie dodatkowych źródeł informacji, to jednak nie zakłóca toku wywiadu właściwego.

Dobrze ułożone winiety, pokrywające różne znaczenia będące w powszechnym w użyciu, pozwalają badaczowi zdobyć wiedzę, jak duża część respondentów pojmuje termin w sposób odpowiadający jego rozumieniu. Tym samym uzyskuje on wgląd, co prawda tylko pośredni, w wielkość i źródła błędów, jakimi obciążone są odpowiedzi na testowane pytanie. Dysponując wynikami, swego rodzaju mapką używanych znaczeń, badacz uzyskuje też pewne wskazówki co do sposobu preredagowania pytania (por. Martin 2006: 4–6).

Ostatnią powszechnie stosowaną techniką, która wprowadza modyfikację do pilotażu konwencjonalnego, jest **respondent debriefing**. Jako źródła informacji wykorzystuje dodatkowe wypowiedzi respondentów, wywołane pytaniami sondującymi, jednak pytania te, podobnie jak w przypadku winiety, zadawane są

po zakończeniu podstawowego wywiadu. Są one przygotowywane przez badacza (zwykle w formie pytań otwartych) i w jednolitej formie zadawane przez ankieterów. Formuła aranżacyjna tego dodatkowego wywiadu przewiduje poinformowanie respondenta o celach badania i pewnej zmianie oczekiwań odnośnie do roli, w jakiej ma on wystąpić – bardziej informatora i kogoś, kto pomaga badaczowi przygotować dobre pytania, niż respondenta (Martin 2006: 11). Charakterystyka ta nasuwa myśl, że polskim odpowiednikiem *respondent debriefing* jest **wywiad o wywiadzie** (Przybyłowska 1975). Jakkolwiek w przypadku obu technik badacze zainteresowani są zdobyciem informacji na temat wewnętrznej reakcji na pytanie kwestionariusza, to w *respondent debriefing* wykorzystuje się je do oceny pytań, natomiast w wywiadzie o wywiadzie – do oceny odpowiedzi. Trzeba podkreślić, że wywiad o wywiadzie nie był skonceptualizowany jako technika badań pilotażowych. Jego zastosowanie (przynajmniej pierwotne) ściśle opierało się na normatywnym modelu reakcji wewnętrznej respondenta na pytanie kwestionariusza autorstwa J. Lutyńskiego, później skonkretyzowanym i dostosowanym do celów badawczych przez B. Tuchańską (1975). Dopasowanie procedury badawczej do modelu i rozbudowanie jej pod tym kątem (Przybyłowska 1975) powoduje, że techniki wywiadu o wywiadzie oraz *respondent debriefing* znacząco się różnią. Oczywiście, nie wyklucza to możliwości wykorzystania elementów tej techniki, co zresztą robili w badaniach pilotażowych K. Lutyńska i F. Sztabiński, do analizy opartej na logice „od oceny odpowiedzi do oceny pytania”.

W kontekście ustaleń zrelacjonowanych przy omówieniu wywiadów kognitywnych, ważną kwestią jest pytanie, czy po upływie nieco dłuższego czasu możliwe jest uzyskanie dostępu do przeżyć, które zaszły pod wpływem pytania kwestionariuszowego. Biorąc pod uwagę ustalenia psychologii poznawczej, uzyskanie bezpośredniego dostępu nie jest możliwe, w tym znaczeniu, że przeżycia te nie stanowią już treści pamięci krótkotrwałej. W takim razie można się zastanowić, na ile możliwe jest ich przypomnienie. W przypadku wywiadu o wywiadzie, pierwotne założenie mówiło, że przypomnienie sobie tych przeżyć i ich werbalizacja są możliwe, natomiast dane empiryczne nie potwierdzały tego w odniesieniu do większości badanych. Mimo wspomagania procesów pamięciowych poprzez odtworzenie nagrania magnetofonowego z pierwszego wywiadu, obserwacje badaczy wskazywały, że większość respondentów werbalizuje swoje aktualne, a nie były przeżycia, niejako formując ponownie odpowiedzi. Jak jednak argumentowano, metoda broni się, gdyż są przesłanki ku temu, by sądzić, że przeżycia aktualne są w dużej mierze tożsame z poprzednimi i stanowią ich powtórzenie oraz kontynuację (Przybyłowska 1975: 468–470). Natomiast podejście badaczy amerykańskich stosujących *respondent debriefing* nie zawsze jest wsparte analizą źródeł informacji i ich stosunku do badanych zjawisk. Za wykorzystaniem tej techniki często stoi tylko ogólne przekonanie o użyteczności wypowiedzi respondentów do dostarczania informacji o wadach

pytań. Jak się wydaje, sprzyja temu fakt, że technika *respondent debriefing* powstała w zasadzie bez teoretycznego zaplecza. Pewne próby odwoływania się do psychologii poznawczej, podejmowane zresztą przez niewielką część badaczy, miały miejsce później.

Oczekiwania w stosunku do tej techniki są na ogół takie, aby ujawniła problemy związane z rozumieniem pytania (np.: Czy mógłby P. własnymi słowami powiedzieć o co chodzi w pytaniu?, Co dla P. znaczy X? Czy X zalicza P. do Y? itp.) czy z drażliwością pytania (np. Czy w wywiadzie było jakieś pytanie, które spowodowało, że poczuł się P. niekomfortowo?) – pod tym względem jej użyteczność jest dobrze oceniana. W badaniu stopnia pewności, z jaką udzielono odpowiedzi, ocena wypada gorzej – respondenci zwykle deklarują silne przekonanie do swoich wcześniejszych wypowiedzi i nie jest to dobry wskaźnik ani problemów z rozumieniem pytania, ani trafności odpowiedzi. Pytania mające na celu odtworzenie toku procesów wewnętrznych są rzadziej stawiane, a ocena informacji, które przynoszą, jest zróżnicowana. Przykład udanego i jednocześnie ambitnego podejścia przedstawia w tym względzie E. Martin, opisując sondaż, w którym pierwotna wersja pytania badającego, czy respondent był ofiarą przestępstwa, dała zaniżone szacunki (37%), co ustalono dzięki wykorzystaniu zewnętrznych źródeł informacji. Nowa wersja pytania (przyniosła wynik 48%) została utworzona w oparciu o ustalenia psychologii poznawczej na temat potencjalnych barier w procesie przypominania i przekazywania informacji. Za pomocą pytań sondujących stwierdzano, czy doszło do poszczególnych zakłóceń. Następnie porównano rezultaty *respondent debriefing* otrzymane dla starej i nowej wersji pytania. W ocenie badaczy uzyskano lepszy wgląd w mechanizm, który stał za różnicą wyników sondażu oraz mocniejsze uzasadnienie do wyboru ostatecznej wersji pytania (Martin 2006: 12–15).

6. Badania eksperymentalne

Badania eksperymentalne stosowane w celu oceny pytań kwestionariusza mogą być dwojakiego rodzaju. Pierwszy rodzaj obejmuje sytuację, w której metodą zbierania danych jest technika pilotażowa, np. wywiad kognitywny. Po randomizacji jednostek, na jednej próbie testuje się oryginalne pytanie, a na drugiej jego nową, ulepszoną wersję, zredagowaną pod wpływem obserwacji problemów, jakie wywołało pytanie w swoim pierwotnym brzmieniu. Sekwencję czynności badawczych można więc przedstawić jako przeprowadzenie pilotażu, a następnie pilotażu po pilotażu. Zastosowanie planu eksperymentalnego ma umożliwić porównywalność wyników poprzez wyeliminowanie czynników wpływu pochodzących od badanej grupy. Metoda nie gwarantuje jednak, że zaobserwowane w fazie testowania różnice w rozkładach odpowiedzi znajdują

odzwierciedlenie w rezultatach badań zasadniczych, co może wynikać z cech próbek (zbyt mała liczebność, niereprezentatywność dla badanej populacji), jak również zależeć od stopnia, w jakim technika pilotażowa odbiega od techniki stosowanej w badaniach właściwych, powodując tym samym różnicę w warunkach udzielania odpowiedzi. Metoda daje jednak możliwość wskazania „lepszego” pytania i dostarczenia uzasadnienia dla takiego wytypowania – „wygrywa” pytanie, które wywołuje mniej problemów, oczywiście tych, których występowanie dana technika pilotażowa jest w stanie ustalić (por. Presser i in. 2004: 118; por. Fowler 2004: 175).

W przypadku drugiego rodzaju badań eksperymentalnych dane zbierane są za pomocą wybranej techniki kwestionariuszowej, typowej dla badań zasadniczych (sondażowych). W jednej z losowych podpróbek stosuje się pytanie w wersji A, w drugiej – w wersji B. Można przypomnieć, że ten typ badań określa się w literaturze anglojęzycznej jako *randomized* lub *split-ballot experiment*. Są one przeprowadzane nie tylko w pilotażowych, ale przede wszystkim metodologicznych celach (przykłady polskich badań: Sułek 2001; Krzewińska 2006) i można je stosować do oceny wpływu wielu innych elementów procedury badawczej, a nie tylko sposobów sformułowania pytania. Wracając jednak do tej kwestii, można powiedzieć, że zaletą *split-ballot experiment* jest możliwość dokładnego sprawdzenia, czy różne wersje pytania generują różne rozkłady odpowiedzi, natomiast jego częstym ograniczeniem jest trudność w wyjaśnieniu tej obserwacji, a co za tym idzie – niemożność wskazania tej wersji pytania, która przynosi bardziej miarodajne rezultaty (Groves i in. 2004: 249–250).

Przezwyciężenie tych trudności wymagałoby od badacza posiadania jakiegoś punktu odniesienia, w świetle którego mógłby zanalizować otrzymane dane. Jeżeli do podstawy, do której odnosi się wyniki, można mieć zaufanie, to różnice w rozkładach odpowiedzi można traktować jako informację o wielkości błędu pomiaru. Najbardziej komfortowym rozwiązaniem byłoby dysponowanie niezależnymi, wiarygodnymi informacjami na temat badanych cech (najlepiej zebranymi na poziomie jednostkowym) i zestawienie ich z uzyskanymi danymi sondażowymi. Jak jednak wiadomo, nieczęsto z takiego rozwiązania udaje się skorzystać, gdyż stosuje się ono jedynie do zmiennych faktualnych, a w praktyce – niewielkiej ich części. Jeżeli nie można skorzystać z zewnętrznych źródeł informacji, podstawą do oceny pytań i zgromadzonych za ich pomocą danych może być koncepcja teoretyczna (zob. Sułek 2001). Czasami, tak jak w przypadku badań mających na celu ustalenie częstości występowania zachowań powszechnie nieakceptowanych, przyjmuje się bardziej zdroworozsądkowe rozumowanie, które tu nakazywałoby wybrać tę wersję pytania, która przynosi wyższe odsetki wskazań (por. Groves i in. 2004: 250). We wszystkich innych sytuacjach pozostaje odwołać się do danych empirycznych, dających się „wewnętrznie” wygenerować. Wart przywołania przykład stanowi propozycja badawcza Fowlera. Zgodnie z nią *split-ballot experiment*, w trakcie którego dokonuje się obserwacji zachowań (*behavior*

coding), powinien być dodatkowo poprzedzony wywiadami kognitywnymi. Wywiady te, zrealizowane na niewielkiej próbie, mają dać wgląd w problemy, jakie stwarzają pierwotne pytania, być podstawą do sformułowania pytań alternatywnych oraz wspomóc analizę wyników eksperymentu. W wyborze ostatecznej wersji pytania należy wziąć dodatkowo pod uwagę wyniki wskaźników behawioralnych informujących o praktycznej użyteczności pytania (*usability*), czyli określających na ile gładko i płynnie, przynajmniej z zewnętrznej perspektywy, przebiega interakcja między ankierem a respondentem w sekwencji pytanie–odpowiedź. Jak unaocznia Fowler, taki schemat badań, choć dość kosztowny, jest efektywny oraz pozwala przezwyciężyć pojawiające się ograniczenia, gdy każda z technik jest stosowana oddzielnie (2004).

7. Podsumowanie

Ten krótki przegląd pokazuje, że zadanie, z którym najlepiej sobie radzą istniejące metody badań pilotażowych, to przede wszystkim identyfikacja problemów stwarzanych przez pytania. Jednocześnie badania nastawione na porównanie efektów, jakie w tym zakresie osiągają różne metody, pokazują, że zgodność jest na ogół mała. Może to świadczyć o tym, że nie ma, przynajmniej na razie, uniwersalnej metody zdolnej diagnozować wszystkie poważne defekty pytania (por. Presser i in. 2004: 124–125).

Jeżeli chodzi o wskazówki odnośnie do sposobu przeredagowania pytania, dają je generalnie te metody, które zakładają wywołanie dodatkowych źródeł informacji. Należy jednak przypomnieć, że jeśli ich zdobycie wymaga zaaranżowania innych warunków udzielania odpowiedzi niż w wywiadzie właściwym, to powstaje wątpliwość, na ile tak uzyskane rezultaty będą się jeszcze do niego odnosić.

Najgorzej natomiast sprawa przedstawia się w odniesieniu do możliwości sprawdzenia, czy pytanie po przeredagowaniu (ten sam problem dotyczy pytań wyjściowych) przynosi miarodajne rezultaty, a przeróbka jest rzeczywiście poprawną. Dzieje się tak na skutek braku możliwości odwołania się do niezależnych źródeł informujących o prawdziwej wartości badanej cechy, standardu – jak określa się to w piśmiennictwie amerykańskim. Niemal wszystkie empiryczne badania, realizowane za pomocą przedstawionych metod, kierowały się logiką typową dla analiz weryfikacji wewnętrznej, zgodnie z którą rezultat ocenia się na podstawie analizy procesu (zob. Lutyński 1975: 331, 362). Nieprawidłowy przebieg procesu formowania informacji (np. niezrozumienie pytania przez duży odsetek respondentów) świadczy negatywnie o wartości zgromadzonego materiału. Na tej samej zasadzie zakłada się, że stwierdzenie braku zakłóceń uprawnia do uznania pytania za narzędzie miarodajne. Stawianie takich wniosków zawsze jest jednak

obarczone ryzykiem. Z uwagi na to, że dostęp do wiarygodnych zewnętrznych informacji jest mocno ograniczony, bardzo cenne są wszystkie te przedsięwzięcia badawcze, w których nie tylko zestawia się wyniki sondażowe, uzyskane za pomocą różnych wersji pytania, ze „standardem”, ale dokładnie obserwuje się przebieg procesu interrogacji dla pytań prowadzących do miarodajnych oraz niemiarodajnych rezultatów, z myślą o odnalezieniu wskaźników trafności oraz nie-trafności odpowiedzi. Wskaźniki takie byłyby bardzo użyteczne w tych wszystkich, jakże częstych, sytuacjach, kiedy do „standardu” odwołać się nie można. Ponadto metodologiczne studia z wykorzystaniem danych zewnętrznych mogą też pomóc w ustaleniu rzeczywistej wartości poszczególnych metod badań pilotażowych (por. Presser i in. 2004: 126).

Podobnie jak oceniono zaprezentowane metody pod kątem stopnia, w jakim realizują cele postawione przed badaniem pilotażowym, tak można je też dokładniej zanalizować pod kątem przedmiotu ich badania. Mówiąc ogólnikowo, przedmiotem testowania i oceny w przypadku każdej z nich jest narzędzie badawcze, ale stwierdzenie to, w świetle dostarczonych charakterystyk poszczególnych podejść, jest mało precyzyjne. Należy bowiem zauważyć, że zbadanie tego, jak funkcjonuje **narzędzie w procedurze** jest możliwe tylko w przypadku tych metod, które pozwalają symulować warunki właściwego wywiadu. Dzięki temu kwestionariusz można potraktować jako pewną całość tworzącą strukturę, której elementami są pytania. Natomiast metody pozbawione realizmu „docelowej” sytuacji badawczej, wnoszące zasadnicze zmiany do procedury wywiadu standaryzowanego, mogą czynić przedmiotem swojego badania **pojedyncze pytania**. Dodatkowe czynności wymagane od respondenta (żądanie pogłębienia i zanalizowania własnej odpowiedzi zaraz po jej wygłoszeniu) rozbijają strukturę narzędzia. Z zastosowaniem tych metod wiąże się też – warte zbadania – zagrożenie wpływu na zachowania respondenta (np. Czy respondent doszedłby do takiej samej odpowiedzi, gdyby był poproszony o myślenie na głos? Czy sondowanie może zmęczyć respondenta i doprowadzić go do obrania strategii, w której będzie się starał tak odpowiadać na pytania kwestionariuszowe, aby zminimalizować ryzyko dalszego dopytywania?). Bynajmniej nie znaczy to, że problem, czy pilotaż powinien mieć charakter symulacyjny, jest rozstrzygnięty i prosty. Zgodnie z tym, co starałam się pokazać w artykule, jest to dylemat, czy przystać na informacje niepełne czy też w miarę pełne, ale niepewne. Dylemat ten staje się szczególnie ostry w świetle ustaleń mówiących, że dostęp do procesów poznawczych istnieje w czasie ich zachodzenia lub krótko po nim, co wymusza ingerencję w procedurę wywiadu standaryzowanego.

Na koniec można też wrócić do pilotażu pogłębionego i zanalizować to podejście w porównaniu z pozostałymi przedstawionymi metodami, z których duża część została opracowana później. Na tym tle pilotaż pogłębiony wykorzystuje dużo źródeł informacji, co stanowi oczywistą zaletę metody, niemniej czyni ją też wymagającą i dość trudną do realizacji w terenie, szczególnie jeżeli zbieranie

tych informacji jest w całości przerzucone na ankietera. Pilotaż pogłębiony korzysta z techniki obserwacji, z tym że w jej jakościowej, niekontrolowanej wersji. Ponadto pod wieloma względami odpowiada technice sondowania (w wywiadzie kognitywnym), osiągając w tym zakresie podobne rezultaty, mimo że podstawa teoretyczna pilotażu pogłębionego oraz techniki sondowania była nieco inna. Biorąc to pod uwagę, pilotaż pogłębiony z pewnością nie „zdezaktualizował się” i jest metodą użyteczną, przynajmniej dla praktyk terenowych. Być może jego zastosowanie w badaniach metodologicznych wymagałoby pewnych modyfikacji, np. nagrywania przebiegu wywiadu w celu prowadzenia obserwacji w bardziej systematyczny sposób.

W artykule nie sposób przedstawić wszystkich stosowanych metod i sposobów mających podnieść wartość narzędzia. Nie omówiono całej grupy technik analiz statystycznych (zob. Presser i in. 2004: 119–121). Nie wspomniano także o innych ciekawych rozwiązaniach, jakimi są różne narzędzia, w tym programy komputerowe, pozwalające na analizę pytania pod kątem jego potencjalnych wad (bez testowania w terenie). QAS (*Question Appraisal System*) (Willis, Lessler 1999) czy QUAID (*Question Understanding Aid*) (Graesser i in. 2006) są dobrymi przykładami takich narzędzi. Bardzo pożądane byłoby, aby polscy badacze mogli korzystać z własnej wersji takich programów, uzyskując coś w rodzaju „pierwszej pomocy przed-pilotażowej”.

Przed metodologami otwiera się też nowe pole badań nad metodami, które pozwolą efektywnie pilotować narzędzia do badań internetowych i innych nowoczesnych form ankiet prowadzonych przy wspomaganiu komputerowym (CASI). Metody takie będą musiały uwzględniać nie tylko pytania kwestionariuszowe, ale także zwracać szczególną uwagę na rozwiązania dotyczące grafiki, obsługi programu itd. Przyrost technik otrzymywania materiałów będzie nasilać potrzebę rozwiązań w tym zakresie.

Ostatecznie zainteresowanie metodologów problematyką badań pilotażowych jest zasadne też z innego powodu – jego brak nigdy nie służy jakości praktyk terenowych.

Bibliografia

- Babbie E. (2003), *Badania społeczne w praktyce*, Warszawa: PWN
- Bassili J. N., Scott B. S. (1996), *Response latency as a signal to question problem in survey research*, „Public Opinion Quarterly”, vol. 60, No. 3
- Beatty P. C., Willis G. B. (2007), *Research synthesis: the practice of cognitive interviewing*, „Public Opinion Quarterly”, vol. 71, No. 2
- Conrad F., Blair J. (1996), *From impressions to data: increasing the objectivity of cognitive interviews*, Proceedings of the 1996 Section on Survey Research Methods, American Statistical Association

- Conrad F., Blair J., Tracy E. (2000), *Verbal reports are data! A theoretical approach to cognitive interviews*, Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference, Washington
- Daniłowicz P. (1974), *Próba standaryzacji zachowań ankierskich w badaniach opartych o wywiad kwestionariuszowy*, [w:] Gostkowski Z. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 3, Warszawa: IFiS PAN
- Daniłowicz P. (1975), *Refleksje nad zastosowaniem pytań pustych w wywiadach kwestionariuszowych*, [w:] Gostkowski Z., Lutyński J. (red.) *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Daniłowicz P., Sztabiński P. B. (1974), *Dwa typy dyrektyw w instrukcjach pilotażowych i konsekwencje ich stosowania*, [w:] Gostkowski Z. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 3, Warszawa: IFiS PAN
- DeMaio T., Ciochetto S., Davis W. (1993), *Research on continuing survey of food intakes by individuals*, Proceedings of the 1993 Section on Survey Research Methods, American Statistical Association
- Draisma S., Dijkstra W. (2004), *Response latency and (para)linguistic expressions as indicators of response error*, [w:] Presser S., Rothgeb J. M., Couper M. P., Lessler J. T., Martin E., Martin J., Singer E. (eds.), *Methods for testing and evaluating survey questionnaires*, New York: John Wiley & Sons
- Esposito J. L., Campanelli P. C., Rothgeb J., Polivka A. E. (1991), *Determining which questions are best: methodologies for evaluating survey questions*, Proceedings of the 1991 Section on Survey Research Methods, American Statistical Association
- Foddy W. (2002), *The in-depth testing of survey questions: a critical appraisal of methods*, Vaus de D. (ed.), *Social surveys*, vol. 3, London: Sage Publications (przedruk artykułu z “Quality and Quantity”, vol. 30, No. 4)
- Fowler F. J., Jr. (1992), *How unclear terms affect survey data*, “Public Opinion Quarterly”, vol. 56, No. 2
- Fowler F. J., Jr. (2004), *The case for more split-sample experiments in developing survey instruments*, [w:] Presser S., Rothgeb J. M., Couper M. P., Lessler J. T., Martin E., Martin J., Singer E. (eds.), *Methods for testing and evaluating survey questionnaires*, New York: John Wiley & Sons
- Gostkowski Z. (1975), *Wstęp*, [w:] Gostkowski Z., Lutyński J. (red.) *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Graesser A. C., Cai Zhiqiang, Louwerse M. M., Daniel F. (2006), *Question Understanding Aid (QUAID). A web facility that tests question comprehensibility*, “Public Opinion Quarterly”, vol. 70., No. 1
- Groves R. M., Fowler F. J., Jr, Couper M. P., Lepkowski J. M., Singer E., Tourangeau R. (2004), *Survey methodology*, New York: John Wiley & Sons
- Krzewińska A. (2006), *O socjologicznym wywiadzie kwestionariuszowym: psychologia poznawcza Norberta Schwarza i jej badawcze zastosowanie w warunkach polskich*, [w:] Gostkowski Z., Kubiak A. (red.), *Analizy i próby technik badawczych w socjologii*, t. 11, Łódź: Łódzkie Towarzystwo Naukowe
- Loftus E. F. (1984), *Protocol analysis of response to survey recall questions*, [w:] Jabine T. B., Straf M. L., Tanur J. M., Tourangeau R. (eds.), *Cognitive aspects of survey methodology. Building a bridge between disciplines*, Washington: National Academy Press
- Lutyńska K. (1974), *Analiza istotności odpowiedzi uzyskiwanych w pogłębionym badaniu pilotażowym*, [w:] Gostkowski Z. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 3, Warszawa: IFiS PAN

- Lutyńska K. (1975), *Pilotaż „pogłębiany”. Koncepcja, realizacja i analiza materiałów pilotażowych*, [w:] Gostkowski Z., Lutyński J. (red.), *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Lutyńska K. (1975a), *Weryfikacja danych uzyskanych z odpowiedzi respondentów na pytania o wysokość zarobków*, [w:] Gostkowski Z., Lutyński J. (red.), *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Lutyńska K. (1978), *Drażliwość i trudność pytań metryczkowych w wywiadzie socjologicznym w świetle opinii respondentów i obserwacji. Raport z pilotażu pogłębianego*, [w:] Gostkowski Z. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 5, Warszawa: IFiS PAN
- Lutyńska K. (1984), *Wywiad kwestionariuszowy. Przygotowanie i sprawdzanie narzędzia badawczego*, Wrocław: Ossolineum
- Lutyńska K. (1981), *Uwagi wstępne. Od metodologii „empirycznej” do prób porządkowania pojęć z zakresu metodologii i teorii socjologii*, [w:] Lutyńska K. (red.), *Szkice metodologiczne*, Warszawa: IFiS PAN
- Lutyński J. (1975), *Analizy weryfikacyjne w badaniach z zastosowaniem wywiadu kwestionariuszowego, ich rodzaje i możliwości*, [w:] Gostkowski Z., Lutyński J. (red.), *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Lutyński J. (1979), *Pytanie jako narzędzie w surveyowych badaniach socjologicznych*, „Przegląd Socjologiczny”, nr 2
- Martin E. (2006), *Vignettes and respondent debriefing for questionnaire design and evaluation*, Research Report Series, Survey Methodology #2006-8, strona internetowa <https://www.census.gov/srd/papers/pdf/rsm2006-08.pdf> (dostęp: lipiec 2012)
- Nuckols R. C. (1953), *A note on pre-testing public opinion questions*, „Journal of Applied Psychology”, vol. 37, No. 2
- Oksenberg L., Cannell Ch., Kalton G. (1991), *New strategies for pretesting survey questions*, „Journal of Official Statistics”, vol. 7, No. 3
- Presser S., Couper M. P., Lessler J. T., Martin E., Martin J., Rothgeb J. M., Singer E. (2004), *Methods for pretesting and evaluating survey questions*, „Public Opinion Quarterly”, vol. 68, No. 1
- Przybyłowska I. (1975), *Wywiad o wywiadzie jako technika otrzymywania informacji o wewnętrznej reakcji respondenta na pytanie kwestionariuszowe*, [w:] Gostkowski Z., Lutyński J. (red.), *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Schaeffer N. C., Dykema J., Garbarski D., Maynard D. W. (2008), *Verbal and paralinguistic behaviors in cognitive assessments in survey interview*, Proceedings of the 2008 Section on Survey Research Methods, American Statistical Association
- Sulek A. (2001), *Sondaż polski*, Warszawa: IFiS PAN
- Sztabiński F. (1982), *Ankieta i jej pilotaż*, [w:] Gostkowski Z. (red.), *Z metodologii i metodyki socjologicznych badań terenowych*, z. 7, Warszawa: IFiS PAN
- Sztabiński P. B., Sztabiński F. (2005), *Jak połączyć pilotaż z badaniem próbnym? Przykład Europejskiego Sondażu Społecznego 2004*, „ASK”, nr 14
- Tuchańska B. (1975), *Analiza modelu wewnętrznej reakcji respondenta na pytanie w wywiadzie kwestionariuszowym*, [w:] Gostkowski Z., Lutyński J. (red.) *Analizy i próby technik badawczych w socjologii*, t. 5, Wrocław–Warszawa–Kraków–Gdańsk: Ossolineum
- Van der Couwen J., Smit J. H. (2004), *Evaluating survey questions by analyzing patterns of behavior codes and questions-answer sequences: a diagnostic approach*, [w:] Presser S., Rothgeb J. M., Couper M. P., Lessler J. T., Martin E., Martin J., Singer E. (eds.), *Methods for testing and evaluating survey questionnaires*, New York: John Wiley & Sons

- Van Someren M. W., Bernard Y. F., Sanberg A.C. (1994), *The think aloud method. A practical guide to modeling cognitive processes*, London: Academic Press
- Willis G. B. (1999), *Cognitive interviewing: A "How To" guide*, Meeting of the 1999 American Statistical Association, Research Triangle Institute
- Willis G. B., Lessler J. T. (1999), *Question appraisal system. QUS-99*, Research Triangle Institute, <http://appliedresearch.cancer.gov/areas/cognitive/qas99.pdf>, (dostęp: wrzesień 2012)

Katarzyna Grzeszkiewicz-Radulska

METHODS FOR PRETESTING SURVEY QUESTIONS

Summary: The tradition of Polish research on pretesting methods refers to “Lodz Methodological School” and dates back to the 70s and 80s. Currently, the interest in testing and evaluating research tools is typical mainly of US methodological literature. The article presents the achievements and outlines the most important field methods of pretesting survey questions. The description of each method contains the review of research procedures, sampling techniques and data analysis in particular, as well as comments on possibilities of practical application. Additionally, we analyse the sources of information used in different methods, as well as point out problems concerning the simulation of research conditions characteristic of “real” interview. Moreover, we sum up the presented methods, paying special attention to the fact to what extent they achieve the goals of pretesting research.

Key words: conventional pretesting, in-depth pilot study, cognitive interviewing, behavior coding, response latency, vignettes, respondent debriefing, experiment.