

31 maja 2014

Przegląd Socjologii Jakościowej

Tom X
Numer 2

*Analiza danych
jakościowych wspomagana
komputerowo*

pod redakcją
Jakuba Niedbalskiego



Polskie Towarzystwo Socjologiczne

www.przegladsocjologiijakosciowej.org

REDAKTOR NACZELNY

Krzysztof Tomasz Konecki, UŁ

REDAKTORZY PROWADZĄCY

Jakub Niedbalski, UŁ

Izabela Ślęzak, UŁ

REDAKTORZY TEMATYCZNI

Waldemar Dymarczyk, UŁ

Marek Gorzko, US

Anna Kacperczyk, UŁ

Sławomir Magala, Erasmus

University

Łukasz T. Marciniak, UŁ

REDAKTOR

DZIAŁU RECENZJI

Dominika Byczkowska, UŁ

REDAKTOR JĘZYKOWY

Aleksandra Chudzik (j. polski)

Jonathan Lilly (j. angielski)

REDAKTOR STATYSTYCZNY

Piotr Chomczyński, UŁ

AUDYTOR WEWNĘTRZNY

Anna Kubczak, UŁ

KOREKTA

Magdalena Chudzik

Magdalena Wojciechowska, UŁ

SKŁAD

Magdalena Chudzik

PROJEKT OKŁADKI

Anna Kacperczyk, UŁ

WWW

Edyta Mianowska, UZ

ADRES REDAKCJI

Uniwersytet Łódzki

Wydział Ekonomiczno-Socjologiczny

Instytut Socjologii

Katedra Socjologii

Organizacji i Zarządzania

ul. Rewolucji 1905 r. 41/43

90-214 Łódź

redakcja.psj@gmail.com

Przegląd
Socjologii
Jakościowej

Tom X
Numer 2

*Analiza danych jakościowych
wspomagana komputerowo*

pod redakcją
Jakuba Niedbalskiego



Polskie Towarzystwo Socjologiczne

Publikacja ukazuje się w ramach Serii Wydawniczej

Polskiego Towarzystwa Socjologicznego,

Edycja IV – „Co po kryzysie?”.

Publikacja dofinansowana została ze środków PTS.

Rada Wydawnicza Serii: Krzysztof Podemski (przewodniczący),

Krzysztof Konecki, Jarosław Kiliński, Olga Nowaczyk

Prawa autorskie

Czasopismo oraz wszystkie zamieszczone w nim artykuły stanowią dorobek współczesnej socjologii. Mogą zostać wykorzystane bez specjalnej zgody dla celów naukowych, edukacyjnych, poznawczych i niekomercyjnych z podaniem źródła, z którego zostały zaczerpnięte.

Wykorzystywanie ogólnodostępnych zasobów zawartych w naszym piśmie dla celów komercyjnych lub marketingowych wymaga uzyskania specjalnej zgody od wydawcy. Pobieranie opłat za dostęp do informacji lub artykułów zawartych w naszym piśmie lub jakiegokolwiek ograniczanie do niego dostępu jest zabronione. Autorzy nadsyłanych artykułów ponoszą odpowiedzialność za uzyskanie zezwoleń na publikowanie materiałów, do których prawa autorskie są w posiadaniu osób trzecich.

Logotyp, szata graficzna strony oraz nazwa *Przeglądu Socjologii Jakościowej* (*Qualitative Sociology Review*) znajdują się w wyłącznym posiadaniu wydawcy. Wszystkie pozostałe obiekty graficzne, znaki handlowe, nazwy czy logotypy zamieszczone na tej stronie stanowią własności ich poszczególnych posiadaczy.

RADA NAUKOWA

Jan K. Coetzee

University of the Free State, Bloemfontein, South Africa

Markieta Domecka

University of Surrey, UK

Aleksandra Galasińska

University of Wolverhampton, UK

Piotr Gliński

Uniwersytet Białostocki

Marek Kamiński

New York University, USA

Michał Krzyżanowski

Lancaster University, UK

Anna Matuchniak-Krasuska

Uniwersytet Łódzki

Barbara Misztal

University of Leicester, UK

Janusz Mucha

Akademia Górniczo-Hutnicza

Sławomir Partycki

Katolicki Uniwersytet Lubelski

Anssi Perakyla

University of Helsinki, Finland

Robert Prus

University of Waterloo, Canada

Marek Szczepański

Uniwersytet Śląski

Piotr Sztompka

Uniwersytet Jagielloński

SPIS TREŚCI

Od redaktora

Jakub Niedbalski

CAQDAS – nowoczesne narzędzia wspomaganie komputerowego w badaniach jakościowych 6

Artykuły

Grzegorz Bryda

CAQDAS a badania jakościowe w praktyce 12

Maciej Brosz

Badanie jakościowe w warunkach niedoszacowania czasu – czyli o konieczności „podążania na skróty”. Organizacja procesu badawczego pod kątem zastosowania komputerowego wsparcia analizy danych jakościowych 40

Jakub Niedbalski

Zastosowanie oprogramowania Atlas.ti i NVivo w realizacji badań opartych na metodologii teorii ugruntowanej 60

Jacek Burski, Kamil Brzeziński

Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” 82

Victoria Kamasa

Techniki językoznawstwa korpusowego wykorzystywane w krytycznej analizie dyskursu. Przegląd 100

Krzysztof Tomanek

Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych 118

Jan Winczorek

Wykorzystanie oprogramowania R i RQDA w jakościowo-ilościowej analizie treści orzeczeń Trybunału Konstytucyjnego 138

Kamil Głowacki

Organizacja procesu badawczego a oprogramowanie do organizacji wiedzy i zarządzania projektem badawczym 162

Artur Piszek, Krzysztof Stachura

Evernote: zastosowanie notatnika internetowego do badań jakościowych 196

Recenzje

Piotr Siuda

Recenzja książki: Magdalena Kamińska (2011) *Niechne memy. Dwanaście wykładów o kulturze Internetu*. Poznań: Galeria Miejska Arsenał 212

Kazimierz Kowalewicz

Recenzja książki: Tomasz Ferenc (2012) *Artysta jako Obcy. Socjologiczne studium artystów polskich na emigracji*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, Wydawnictwo Biblioteki Państwowej Wyższej Szkoły Filmowej, Telewizyjnej i Teatralnej 218

Jakub Niedbalski
Uniwersytet Łódzki

DOI: <https://doi.org/10.18778/1733-8069.10.2.01>

Od redaktora:

CAQDAS – nowoczesne narzędzia wspomaganie komputerowego w badaniach jakościowych

Abstrakt Już od dłuższego czasu obserwuje się rosnące zainteresowanie badaczy, naukowców, ale także praktyków prowadzących badania jakościowe narzędziami, które mogłyby wspomóc proces analityczny. Mimo to tematyka CAQDAS (komputerowego wspomaganie analizy danych jakościowych) nie znalazła jak dotąd swojego miejsca w zbyt wielu rodzimych opracowaniach. To skłoniło nas do podjęcia współpracy w szerszym gronie badaczy, czego wynikiem jest ów tom, poświęcony niniejszej problematyce. Znajdują się w nim teksty osób, które wzięły udział w grupie tematycznej zorganizowanej podczas XV Zjazdu Socjologicznego w Szczecinie, ale także artykuły innych badaczy posiadających doświadczenie w stosowaniu nowoczesnych narzędzi wspomagających proces badawczy. Proponowany zbiór artykułów ma szansę stać się jedną z pierwszych tego typu publikacji, która zaprezentuje możliwości i sposób wykorzystania różnych programów CAQDAS w badaniach opartych na metodach jakościowych.

Słowa kluczowe CAQDAS, komputerowe wspomaganie analizy danych jakościowych, metody badań społecznych, metody badań jakościowych, nowe technologie

Jakub Niedbalski, doktor socjologii, adiunkt w Katedrze Socjologii Organizacji i Zarządzania Instytutu Socjologii Uniwersytetu Łódzkiego. Specjalizuje się w komputerowej analizie danych jakościowych, metodach badań jakościowych, zagadnieniach socjologii niepełnosprawności i socjologii kultury fizycznej. Prowadzi badania poświęcone aktywizacji społecznej i fizycznej osób z niepełnosprawnością. Autor książek: *Życie i praca w domu pomocy społecznej. Socjologiczne studium interakcji personelu z upośledzonymi umysłowo podopiecznymi* (Wydawnictwo UŁ, Łódź 2013), *Odskrywanie CAQDAS. Wybrane bezpłatne programy komputerowe*

wspomagające analizę danych jakościowych (Wydawnictwo UŁ, Łódź 2013) oraz *Komputerowe wspomaganie analizy danych jakościowych. Zastosowanie oprogramowania NVivo i Atlas.ti w projektach badawczych opartych na metodologii teorii ugruntowanej* (Wydawnictwo UŁ, Łódź 2014).

Adres kontaktowy:

Katedra Socjologii Organizacji i Zarządzania
Instytut Socjologii Uniwersytetu Łódzkiego
ul. Rewolucji 1905 r. 41, 90-214 Łódź
e-mail: jakub.niedbalski@gmail.com

Nowe technologie w istotny sposób przeobrażają proces prowadzenia i realizacji badań naukowych. Dotyczy to nie tylko dyscyplin ścisłych czy przyrodniczych, ale też społecznych i humanistycznych (Hammersley, Atkinson 2000; Babbie 2003; Silverman 2008; Denzin, Lincoln 2009; Kubinowski 2010). Jednym z takich kierunków jest dynamiczny rozwój oprogramowania wspomagającego analizę danych jakościowych, który na świecie widoczny jest od przeszło dwóch dekad. W ciągu tego czasu, wychodząc naprzeciw środowisku naukowemu, powstało wiele rozmaitych narzędzi komputerowych, które stawały się coraz doskonalsze, oferując ich użytkownikom kolejne zaawansowane funkcje. Znajduje to także potwierdzenie w stale rosnącej ofercie oraz możliwościach oprogramowania CAQDAS (*Computer-Assisted Qualitative Data Analysis Software*) (Niedbalski, Ślęzak 2012).

Niniejsza publikacja inspirowana jest aktualnymi trendami w naukach społecznych i humanistycznych na świecie. Również w Polsce mamy do czynienia z rosnącym zainteresowaniem zarówno świata akademickiego, jak też przedstawicieli firm i instytucji państwowych, które coraz częściej sięgają po różnego rodzaju narzędzia komputerowego wspomaganie analizy danych jakościowych. Ale mimo że rośnie liczba badaczy, naukowców, a także praktyków zainteresowanych prowadzeniem badań jakościowych, poszukujących przy tym narzędzi, które mogłyby wspomóc proces analityczny, nadal powstaje niewiele publikacji na ten temat. Warto przy tym podkreślić, że osoby zainteresowane metodami ilościowymi mogą czerpać z bogatej literatury prezentującej takie programy komputerowe, jak SPSS czy Statistica, a także wiele innych.

Natomiast na rodzimym rynku wydawniczym nie ma jak do tej pory zbyt wielu publikacji odnoszących się do programów CAQDAS. W polskiej literaturze metodologicznej występują jedynie pojedyncze opracowania dotyczące tego tematu (Trutkowski 1999; Bieliński, Iwińska, Rosińska-Kordasiewicz 2007; Niedbalski, Ślęzak 2012; Brosz 2012; Niedbalski 2013; 2014). Dlatego powiększanie się grona osób zarówno korzystających z oprogramowania komputerowego, jak i zainteresowanych jego wdrożeniem w planowanych i realizowanych przez siebie przedsięwzięciach badawczych, które jak dotąd nie miały okazji do wymiany doświadczeń, prezentacji swoich dokonań, a z drugiej strony – uzyskania szerokiej i fachowej wiedzy w zakresie możliwości czy sposobów zastosowania tego rodzaju narzędzi informatycznych, stało się bezpośrednim powodem wydania tego tomu. Niniejsza publikacja ma za zadanie przybliżyć nowe spojrzenie na metodologię badań jakościowych i przyczynić się do lepszego zrozumienia stojących za nią idei, zwłaszcza wśród przedstawicieli nauk społecznych i humanistycznych, a także uzupełnić lukę, jaka do tej pory istniała (poza nielicznymi wyjątkami) na gruncie polskiej literatury dotyczącej komputerowego wspomaganie analizy danych jakościowych.

Prezentowany tom zawiera teksty przygotowane przez badaczy i praktyków, dla których praca w środowisku oprogramowania komputerowego jest codziennością. Dzięki temu otrzymujemy bardzo rzetelną wiedzę opartą na wieloletnim doświadczeniu poszczególnych autorów, którzy w danym temacie posiadają wiedzę ekspercką. Publikacja jest zbiorem tekstów, które w sposób przekrojowy pokazują różnorodność i specyfikę programów komputerowych

wspomagających analizę danych jakościowych, a także prezentują ich wykorzystanie w ramach badań opartych na rozmaitych metodach badawczych i prowadzonych z wykorzystaniem różnych rodzajach danych. W ten sposób zyskujemy szerokie spektrum informacji oraz wiedzy co do możliwości aplikacyjnych popularnych programów należących do rodziny CAQDAS. Wydaje się zatem, że prezentowana publikacja powinna zarówno zaspokoić oczekiwania niedoświadczonych jeszcze użytkowników oprogramowania, którzy pragną zobaczyć, na czym polega korzystanie z tego rodzaju oprogramowania, jak i znaleźć odbiorców wśród bardziej wytrawnych badaczy, którzy dzięki niej mogą zrewidować swój warsztat badawczy, a być może odnaleźć źródło dodatkowych inspiracji.

Wśród wielu zagadnień, jakim autorzy poświęcili swoje teksty, można wyróżnić te, w których podejmuje się dyskusje nad zgodnością zasad, na jakich funkcjonuje oprogramowanie CAQDAS, z regułami oraz procedurami metodologii badań jakościowych, a także możliwościami zastosowania oprogramowania CAQDAS w realizacji projektów badawczych opartych na różnych metodach jakościowych i w ramach różnych podejść analitycznych. W prezentowanym tomie nie zabrakło również tekstów pokazujących wpływ nowych technologii na przebieg procesu badawczego czy podejmujących refleksję nad kierunkami rozwoju nowoczesnych programów komputerowych, a także nad przyszłością metod jakościowych w kontekście zastosowania oprogramowania CAQDAS.

Niniejszy numer otwiera tekst Grzegorza Brydy „CAQDAS a badania jakościowe w praktyce”. Na

podstawie długoletniego doświadczenia w korzystaniu z różnych narzędzi CAQDAS autor stara się ukazać ich wpływ na zmianę sposobu myślenia o metodologii, procesie analizy danych i prowadzeniu terenowych badań jakościowych. Według Grzegorza Brydy korzystanie z oprogramowania komputerowego kształtuje osobowość i tożsamość badacza jakościowego, jego styl pracy i warsztat analityczno-badawczy oraz sprzyja rozwijaniu nowych umiejętności analitycznych i informatycznych, bez których, zdaniem autora, trudno wyobrazić sobie współczesne jakościowe badania socjologiczne.

W kolejnym artykule zatytułowanym „Badanie jakościowe, na które nie przewidziano czasu – czyli o konieczności «podążania na skróty»». Organizacja procesu badawczego pod kątem zastosowania komputerowego wsparcia analizy danych jakościowych” Maciej Brosz prezentuje zastosowanie programu NVivo jako kluczowego elementu warsztatu metodologicznego w badaniach jakościowych, prowadzonych w warunkach wąskich ram czasowych. Zdaniem autora tekstu nowoczesne technologie, takie jak specjalistyczne oprogramowanie komputerowe, znajdujące wyraz w najnowszych wydaniach specjalistycznego oprogramowania, dostarczają takiej funkcjonalności, która – w połączeniu z rosnącymi mocami obliczeniowymi komputerów – umożliwia stawianie czoła coraz to nowym wymaganiom współczesności.

Również tekst Jakuba Niedbalskiego „Zastosowanie oprogramowania Atlas.ti i NVivo w realizacji badań opartych na metodologii teorii ugruntowanej” porusza problematykę zastosowania specjalistyczne-

go oprogramowania wspomagającego analizę danych jakościowych, przy czym w tym wypadku autor czyni to w kontekście realizacji badań opartych na procedurach metodologii teorii ugruntowanej (MTU). Dlatego celem artykułu jest pokazanie, jakie związki istnieją pomiędzy procedurami metodologii teorii ugruntowanej a dwoma popularnymi programami z rodziny CAQDAS: NVivo oraz Atlas.ti.

Ze względu na podkreślenie aplikacyjnego wymiaru zastosowania oprogramowania komputerowego w badaniach jakościowych podobieństwo do wcześniejszych artykułów wykazuje tekst Jacka Burskiego i Kamila Brzezińskiego. Dotyczy on wykorzystania programu komputerowego QDAMiner do realizacji konkretnego projektu badawczego. Celem, jaki stawiają sobie w artykule jego autorzy, jest przedstawienie podstawowych założeń metodologicznych i teoretycznych określonego projektu badawczego.

Kolejny tekst, którego autorką jest Victoria Kamasa – „Metody korpusowe w jakościowej analizie tekstu na przykładzie krytycznej analizie dyskursu” – koncentruje się na analizie roli dyskursu w umacnianiu i reprodukowaniu relacji władzy i dominacji. Na tej podstawie autorka kreśli rolę oprogramowania komputerowego służącego strategii doboru korpusu do badań w ramach KAD, a także jego wykorzystanie w realizacji takich metod opracowywania materiału, jak analiza frekwencji, analiza słów kluczowych, analiza kolokacji i analiza konkordancji.

Krzysztof Tomanek w artykule „«Analiza sentymentu» – metoda analizy danych jakościowych. Przykład zastosowania i ewaluacja słownika RID i Bayesa metody klasyfikacji w analizie danych jakościowych”

podkreśla z kolei znaczenie oprogramowania komputerowego wspierającego metody klasyfikacji jakościowych danych tekstowych. Autor zwraca przede wszystkim uwagę na rozwiązania mające na celu zbudowanie trafnego słownika, w kontekście klasyfikacji tekstów. W tym celu porównuje skuteczność analiz nadzorowanych do skuteczności analiz zautomatyzowanych, dochodząc przy tym do wniosku, że słownik, który w przeszłości uzyskał dobrą ocenę jako narzędzie klasyfikacyjne, wtedy, gdy stosowany jest wobec nowego materiału empirycznego, powinien zostać poddany weryfikacji i procedurom modyfikacji.

Kolejny artykuł autorstwa Jana Winczorka pokazuje łączenie w jednym projekcie badawczym jakościowych i ilościowych metod badawczych, przy wspieraniu się w takim przedsięwzięciu określonymi narzędziami komputerowej analizy danych. Autor opisuje wykorzystanie pakietu statystycznego „R” z nakładką CAQDA „RQDA” w jakościowo-ilościowej triangulowanej analizie treści z kodowaniem swobodnym, analizą rzetelności kodowania i statystyczną istotnością wyników, przeprowadzonej na próbie uzasadnień orzeczeń Trybunału Konstytucyjnego z lat 1986–2009.

Numer zamykają dwa artykuły poświęcone prezentacji niezwykle pożytecznych narzędzi wspomagających pracę badacza w organizacji jego warsztatu. Artykuł Kamila Głowackiego zawiera obszerne zestawienie najbardziej popularnych, a zarazem uniwersalnych programów komputerowych, które można wykorzystać na różnych etapach procesu badawczego, począwszy od wstępnej organizacji danych, poprzez właściwą analizę, a na formatowaniu

końcowego jej produktu w postaci publikacji skończywszy. Z kolei Artur Piszek i Krzysztof Stachura prezentują w swoim artykule aplikację Evernote, którą traktują jako prostszą i darmową alternatywę dla profesjonalnych pakietów przeznaczonych do jakościowej analizy danych. Autorzy przedstawiają narzędzia i opcje pozwalające na aplikację metodologii teorii ugruntowanej przy pomocy opisywanego programu oraz wskazują na podobieństwa i różnice pomiędzy prezentowanym oprogramowaniem a pakietem NVivo.

Bibliografia

Babbie Earl (2003) *Badania społeczne w praktyce*. Wyd. pol. pod red. Agnieszki Kloskowskiej-Dudzińskiej. Przełożył zespół tłumaczy. Warszawa: Wydawnictwo Naukowe PWN.

Bieliński Jacek, Iwańska Katarzyna, Rosińska-Kordasiewicz Anna (2007) *Analiza danych jakościowych przy użyciu programów komputerowych*. „ASK. Społeczeństwo. Badania. Metody”, nr 16, s. 89–114.

Brosz Maciej (2012) *Komputerowe wspomaganie badań jakościowych. Zastosowanie pakietu NVivo w analizie materiałów nieustrukturyzowanych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 98–125 [dostęp 20 listopada 2013 r.]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

Denzin Norman K., Lincoln Yvonna S., red., (2009) *Metody badań jakościowych*, t. 1. Wyd. pol. pod red. Krzysztofa Podemskiego. Przełożył zespół tłumaczy. Warszawa: Wydawnictwo Naukowe PWN.

Hammersley Martyn, Atkinson Paul (2000) *Metody badań terenowych*. Przełożył Sławomir Dymczak. Poznań: Zysk i S-ka.

Kubinowski Dariusz (2010) *Jakościowe badania pedagogiczne*. Lublin: Wydawnictwo UMCS.

Wszystkie zawarte w tomie teksty stanowią istotny wkład w zrozumienie specyfiki oraz istoty rozmaitych kontekstów i uwarunkowań związanych ze stosowaniem oprogramowania komputerowego wspomagającego analizę danych jakościowych. Z całą pewnością mogą przyczynić się do lepszego poznania problematyki CAQDAS oraz wzbudzić refleksję nad aktualnym stanem wiedzy dotyczącym dynamicznie rozwijanego na całym świecie specjalistycznego oprogramowania komputerowego.

Niedbalski Jakub (2013) *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub (2014) *Komputerowe wspomaganie analizy danych jakościowych. Zastosowanie oprogramowania Nvivo i Atlas.ti w projektach badawczych opartych na metodologii teorii ugruntowanej*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 126–165 [dostęp 20 listopada 2013 r.]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

Silverman David (2008) *Interpretacja danych jakościowych*. Przełożyła Joanna Ostrowska. Warszawa: Wydawnictwo Naukowe PWN.

Trutkowski Cezary (1999) *Analiza treści wspomagana komputerowo. Badanie społecznych reprezentacji polityki*. „ASK. Społeczeństwo. Badania. Metody”, nr 8, s. 113–133.

Cytowanie

Niedbalski Jakub (2014) *Od redaktora: CAQDAS – nowoczesne narzędzia wspomaganie komputerowego w badaniach jakościowych*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 6–11 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

CAQDAS – Modern Computer-Aided Tools Applied in Qualitative Data Collection

Abstract: For a long time, a growing interest in tools which could support an analytical process has been noticed among researchers, scientists, as well as practitioners who run a qualitative data collection. However, CAQDAS (computer-assisted qualitative data analysis) gained popularity only in several domestic studies. This is the reason why we decided to work within a bigger group of researchers, what resulted in this very volume devoted to the above-mentioned subject. It embraces essays of people who participated in a focus group organized during the 15th Congress of the Polish Sociological Association in Szczecin, as well as articles of other researchers who are experienced in applying modern tools which support research process. This collection of articles may become one of the first such publications which present the possibilities and manners of using various types of CAQDAS programs in research based on qualitative methods.

Keywords: CAQDAS, Computer Assisted Qualitative Data Analysis Software, Social Research Methods, Qualitative Research Methods, New Technologies

Grzegorz Bryda Uniwersytet Jagielloński

DOI: <https://doi.org/10.18778/1733-8069.10.2.02>

CAQDAS a badania jakościowe w praktyce

Abstrakt Celem artykułu jest refleksja metodologiczna nad dialektyczną relacją pomiędzy procesem badawczym a procesem wspomaganą komputerowo analizy danych jakościowych. Na podstawie długoletniego doświadczenia w korzystaniu z różnych narzędzi CAQDAS autor stara się ukazać ich wpływ na zmianę sposobu myślenia o metodologii, procesie analizy danych i prowadzeniu terenowych badań jakościowych. Specyfika użycia CAQDAS w praktyce badawczej wymaga swoistego rygorystycznego metodologicznego w procesie gromadzenia i archiwizacji danych oraz dokładności i precyzji w procesie kodowania, analizy i wizualizacji danych jakościowych. Zastosowanie komputerowej analizy danych jakościowych w praktyce badawczej kształtuje nie tylko ramy interpretacji socjologicznej, ale przede wszystkim zmienia optykę i sposób percepcji problemów badawczych. Istotą tego procesu jest swoista interakcja pomiędzy nowymi technologiami a tradycyjną metodologią analizy danych i prowadzenia badań jakościowych. W tym sensie korzystanie z CAQDAS w praktyce badań jakościowych kształtuje osobowość i tożsamość badacza jakościowego, jego styl pracy i warsztat analityczno-badawczy oraz sprzyja rozwijaniu nowych umiejętności analitycznych i informatycznych, bez których trudno wyobrazić sobie współczesne jakościowe badania socjologiczne.

Słowa kluczowe CAQDAS, technologie informatyczne, analiza danych, badania jakościowe, styl analizy danych, metody mieszane (*mixed methods*)

Grzegorz Bryda, dr, adiunkt w Instytucie Socjologii UJ, Pełnomocnik Rektora UJ ds. ewaluacji jakości systemu kształcenia. Zainteresowania: teoria socjologiczna, kognitywistyka, zastosowanie informatyki i NLP w analizie danych jakościowych (CAQDAS i Text Mining), modelowanie procesów społecznych. Współpracuje z instytucjami publicznymi i prywatnymi w zakresie metodologii badań społecznych oraz analizy danych ilościowych i jakościowych.

Adres kontaktowy:

Instytut Socjologii, Uniwersytet Jagielloński
31-044 Kraków, ul. Grodzka 52
e-mail: grzegorz.bryda@uj.edu.pl

Wstęp. CAQDAS a metody jakościowe

Zastosowanie komputerów w naukach społecznych i początki rozwoju oprogramowania do analizy danych sięgają drugiej połowy lat sześćdziesiątych XX wieku (Brent, Anderson 1990; Tesch 1990). Podobnie idea wspomaganą komputerowo analizy danych jakościowych ma długą tradycję w metodologii badań socjologicznych. Początkowo była ona związana z możliwością wykorzystania technologii informatycznych w analizie treści (Berelson 1952), szczególnie w odniesieniu do tak zwanej kwantyfikowanej analizy treści. W ciągu ostatnich dwóch dekad, wraz

z rozwojem technologii informatycznych na masową skalę, pojawiły się szersze możliwości wykorzystania programów do wspomaganą komputerowo analizy danych w tradycyjnych badaniach jakościowych zorientowanych na technikę wywiadów socjologicznych czy analizę treści (Becker, Gordon, LeBailly 1984; Brent 1984; Gerson 1984; Pfaffenberger 1988). Rozwój programów do wspomaganą komputerowo analizy danych jakościowych oraz ich widoczna obecnie różnorodność mają swoje odzwierciedlenie w indywidualnym podejściu badaczy do procesu analizy i dominującej metodologii badań jakościowych. Pomimo że współcześnie różnice te zacierają się coraz bardziej ze względu na postępującą komercjalizację oprogramowania, to paradygmaty metodologiczne, które stały u podstaw ich rozwoju, nadal są obecne w funkcjonalnościach programów.

Punktem zwrotnym w rozwoju oprogramowania do analizy danych jakościowych było powołanie do życia – w 1994 roku na University of Surrey – „CAQDAS Networking Project”¹. Celem projektu była integracja środowiska badaczy jakościowych przez: dostarczenie informacji, organizowanie szkoleń z zakresu wykorzystywania programów do komputerowej analizy danych jakościowych, tworzenie platformy do debaty dotyczącej kwestii analitycznych, metodologicznych i epistemologicznych wynikających z korzystania z oprogramowania CAQDAS oraz prowadzenie badań socjologicznych dotyczących ich zastosowań².

¹ Zob. Strona projektowa The CAQDAS Networking Project: <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/about/>.

² Konsekwencją powstania projektu CAQDAS są cykliczne międzynarodowe konferencje, seminaria i warsztaty organizowane przez badaczy skupionych wokół University of Surrey i projektu. Konferencje te są doskonałą formą nie tylko konfrontacji dokonań badaczy z różnych krajów we wspomaganą komputerowo analizie danych jakościowych, ale przede

Projekt ten powstał w następstwie konferencji metodologicznej, która odbyła się w 1989 roku na University of Surrey. Konferencja zgromadziła badaczy – pionierów stosujących oprogramowanie CAQDAS w analizie danych jakościowych – oraz zaowocowała publikacją metodologów i deweloperów (Fielding, Lee 1993). Publikacja ta stała się swoistym manifestem środowiska osób zainteresowanych rozwojem oprogramowania i technologii CAQDAS, stymulującym dyskurs metodologiczny nad jego wykorzystaniem w analizie danych jakościowych w obszarze nauk społecznych, a przede wszystkim socjologii. Jednocześnie pomysłodawcy wspomnianego projektu, Raymond Lee i Nigel Fielding, wprowadzili do powszechnego użycia termin CAQDAS (akronim angielskiego terminu *Computer-Assisted Qualitative Data Analysis Software*) lub krócej QDAS (*Qualitative Data Analysis Software*).

Pierwsze programy do komputerowej analizy danych jakościowych były pisane przez badaczy – entuzjastów, którzy nie tylko sami realizowali badania terenowe czy prowadzili analizy, ale posiadali umiejętności programowania lub znali kogoś, kto je miał. Wielu z nich rozwijało programy niezależnie od siebie, często pozostając nieświadomymi faktu, że inni również pracują nad tego typu narzędziami analitycznymi. Brakowało takich możliwości, jakie dziś zapewnia Internet, to jest zróżnicowanej formy wymiany informacji oraz komunikowania się deweloperów między sobą

wszystkim forum spotkań i dyskusji między badaczami, analitykami czy deweloperami. W 2007 roku uczestniczyłem, jako jeden z trójki polskich badaczy jakościowych, w konferencji „CAQDAS: Advances in Qualitative Software” zorganizowanej w Royal Holloway, University of London. Niestety doświadczenie to pokazuje, że środowisko polskich badaczy jakościowych jest słabo reprezentowane w sieci CAQDAS. Zob. http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/trainingandevents/caqdas_2007_conference_advances_in_qualitative_software.htm.

i z użytkownikami³. Nie było regularnych konferencji z zakresu wspomaganej komputerowo analizy danych jakościowych czy warsztatów organizowanych przez deweloperów, forów użytkowników, biuletynów informacyjnych, a przede wszystkim literatury przedmiotu. Programy do komputerowej analizy danych jakościowych były rozwijane od połowy lat osiemdziesiątych – dla komputerów na platformie IBM PC w Stanach Zjednoczonych, Niemczech, Wielkiej Brytanii, Danii, Holandii i Australii. Wraz z pojawieniem się pierwszych programów takich jak Text Base Alpha, Ethno, Qualpro, TAP czy The Ethnograph (Drass 1989; Tesch 1990; Fischer 1994) wykorzystanie komputerów w analizie danych jakościowych wzbudzało szereg kontrowersji wśród badaczy jakościowych.

Na przełomie lat osiemdziesiątych i dziewięćdziesiątych XX wieku w wielu publikacjach naukowych z zakresu socjologii dotyczących wspomaganej komputerowo analizy danych przewijała się debata na temat możliwości oraz pozytywnych i negatywnych skutków zastosowania oprogramowania w metodologii badań jakościowych (Conrad, Reinharz 1984a; Richards, Richards 1989; Richards, Richards 1991; Seidel 1991; Kelle 1995)⁴. Dotyczyła

³ Obecnie praktycznie każdy dostawca oprogramowania CAQDAS, bez względu na to, czy jest ono komercyjne czy nie, korzysta z różnych form komunikacji z użytkownikami. Popularnym rozwiązaniem są kursy e-learningowe oferowane przez dostawców komercyjnych, dotyczące podstaw obsługi programów CAQDAS. Ponadto istnieje szereg forów dyskusyjnych dla użytkowników oprogramowania CAQDAS, na których uczestnicy udzielają sobie porad dotyczących analizy danych jakościowych, konferencji, realizowanych projektów itp.

⁴ Echa tej debaty nadal są, choć w mniejszym stopniu, obecne w dyskursie metodologicznym dotyczącym wykorzystania komputerów w analizie danych jakościowych. Oponenty CAQDAS zwracają uwagę na to, że ciągły rozwój nowych funkcjonalności w programach wpływa negatywnie na zachowanie jakościowego charakteru analiz i badań. Obawy i wątpliwości sceptyków dotyczą przede wszystkim rzetelności metodologicznej prowadzonych analiz CAQDAS i utraty tradycyjnego, jakościowego charakteru danych terenowych.

ona między innymi: wiarygodności analizy, relacji między tradycyjnym podejściem do analizy danych jakościowych a podejściem zorientowanym na CAQDAS, dekontekstualizacji danych jakościowych wynikającej z zapośredniczenia procesu analizy przez wykorzystywanie komputerów, efektywności i zasadności wykorzystania oprogramowania CAQDAS w zależności od różnego rodzaju i zakresu danych. Na podobne problemy zwracają uwagę Raymond Lee i Nigel Fielding (1993; 1996; 2012). Rekonstruując, przy pomocy wywiadów pogłębionych, możliwości i ograniczenia wykorzystywania CAQDAS w procesie badań jakościowych, wskazują na trzy istotne zagadnienia:

- Po pierwsze, stosowanie CAQDAS w badaniach jakościowych powoduje, że program komputerowy staje się „pośrednikiem” w dostępie do materiału empirycznego. Do tej pory badacze prowadzący analizę „na papierze” mieli poczucie bliższej styczności z danymi, na przykład wypowiedziami respondentów, notatkami terenowymi niż analitycy korzystający z CAQDAS. Zdaniem Lee i Fieldinga styczność z materiałem dotyczy lepszej kontekstualizacji analizowanych treści czy wypowiedzi, która to funkcjonalność była słabo rozwinięta we wcześniejszych wersjach programów CAQDAS. Obecnie problem „styczności z danymi” wydaje się być rozwiązywany, z jednej strony, dzięki wielowymiarowym możliwościom pracy z danymi tekstowymi, takimi jak wystandaryzowane etapy i procedury kodowania, stosowanie różnych rodzajów notatek (*memos*) analitycznych i badawczych, hiperlinkowanie tekstów i kodowanych fragmentów w obrębie analizowanego dokumentu lub między do-

kumentami, synchronizacja z danymi źródłowymi audio-wideo, geopozycjonowanie czy analizy profilowe i przekrojowe (Miles, Huberman 2000), zaś z drugiej – poprzez odpowiednie przygotowanie interfejsu danego programu.

- Po drugie, na rozwijanie programów wspomagających jakościową analizę danych i ich funkcjonalności od samego początku miały wpływ dwie metodologie badawcze: teoria ugruntowana (Glaser, Strauss 2009; Bong 2002) i analiza treści (Berelson 1952). Programy były pisane w taki sposób, by ułatwiały prowadzenie analizy danych jakościowych zgodnie z logiką i procedurami wspomnianych metodologii. Lee i Fielding zwracają jednak uwagę, że wraz z rosnącym zaawansowaniem programów do wspomaganej komputerowo analizy danych jakościowych, zmniejsza się jednocześnie ich zależność od konkretnego podejścia metodologicznego.
- Po trzecie, w rozwoju oprogramowania CAQDAS zbyt duży nacisk został położony na procedurę kodowania (*coding*) i wyszukiwania (*retrieval*) treści. Zdecydowana większość programów do wspomaganej komputerowo analizy danych jakościowych opiera się właśnie na tych dwóch podstawowych funkcjonalnościach. Brakuje implementacji nowych technik i procedur analitycznych, co ogranicza w pewnym stopniu możliwości kompleksowego wykorzystywania programów do analizy danych jakościowych w innych obszarach czy dziedzinach nauki⁵.

⁵ Funkcje kodowania i przeszukiwania danych tekstowych są rdzeniem każdego programu do komputerowej analizy danych jakościowych. Warto zwrócić jednak uwagę na to, że

W dalszej części tego artykułu, odnosząc się do powyższych zagadnień oraz wykorzystując wiedzę metodologiczną, doświadczenie w dziedzinie realizacji badań terenowych i pracy z różnymi programami CAQDAS, chciałbym ukazać zalety i wady stosowania tych narzędzi w procesie badań jakościowych. Jednocześnie chciałbym zaznaczyć, że nie jest moim celem dokonywanie porównań w zakresie funkcjonalności pomiędzy poszczególnymi programami CAQDAS, lecz refleksja metodologiczna⁶. Podstawą tych rozważań są projekty i ekspertyzy naukowe, które miałem okazję realizować na przestrzeni ostatnich kilku lat⁷. W latach 2005–2007 kierowałem realizacją dwóch projektów na terenie województwa małopolskiego: ekspertyza efektywności form wsparcia oferowanego długotrwale bezrobotnym w województwie małopolskim (2005–2007)⁸ oraz badanie potrzeb niepełnosprawnych i oczekiwań pracodawców na rynku pracy (2006–2008)⁹. Oba projekty opierały się na analizie danych zastanych i realizacji badań jakościowych. W przypadku pierwszego przeprowadzono 24 grupy

kodowanie i przeszukiwanie stanowią również dwie podstawowe procedury postępowania w tradycyjnej, manualnej analizie danych jakościowych (Dey 1993; Richards 2005).

⁶ Zainteresowanych tego rodzaju porównaniami odsyłam do książki Ann Lewins i Christiny Silver (2007), poświęconej programom komercyjnym, książki Jakuba Niedbalskiego (2013), dotyczącej programów bezpłatnych oraz strony internetowej projektu CAQDAS: http://onlineqda.hud.ac.uk/Step_by_step_software/index.php, która w przystępny sposób wprowadza w świat wspomaganej komputerowo analizy danych jakościowych. Dla cierpliwych godnym polecenia jest zapoznanie się z funkcjonalnościami programów poprzez dostępne na stronach producentów instrukcje.

⁷ Badania jakościowe wspomagane programami CAQDAS realizuję od 1997 roku, jednakże w artykule korzystam z tych doświadczeń projektowych, które w największym stopniu wpłynęły na zmianę mojego sposobu myślenia o zastosowaniu komputerowej analizy danych w badaniach jakościowych.

⁸ Zob. Bryda, Anacik, Jaśko (2007).

⁹ Zob. Bryda, Anacik, Jaśko (2008).

fokusowe, a w przypadku drugiego – 2 razy po 100 wywiadów pogłębionych z osobami niepełnosprawnymi i pracodawcami lokalnych rynków pracy. W latach 2008–2010 kierowałem, w ramach Programu Operacyjnego Kapitał Ludzki, Działanie 5.4 – Rozwój potencjału trzeciego sektora, Poddziałanie 5.4.2 – Rozwój dialogu obywatelskiego, projektem „Diagnoza stanu współpracy organizacji pozarządowych w województwie małopolskim – badania jakościowe”¹⁰. W ramach projektu przeprowadzonych zostało 150 wywiadów pogłębionych z przedstawicielami organizacji pozarządowych i 14 wywiadów grupowych z członkami różnych organizacji na terenie województwa małopolskiego. W latach 2007–2013 realizowałem własny projekt – eksperyment metodologiczny – dotyczący wpływu atrybucji społecznych na proces kształtowania się schematów poznawczych aktorów, ich codziennych praktyk i strategii działania, na przykładzie zbiorowości lokalnej. W ramach projektu przeprowadziłem serię wywiadów biograficznych i swobodnych z mieszkańcami gminy Komańcza położonej w powiecie sanockim na terenie województwa podkarpackiego. Badania zostały zrealizowane przez autora i studentów socjologii w cyklu 10 wyjazdowych obozów naukowo-badawczych jako element kursu Badania Jakościowe w Praktyce (Warsztaty z Analizy i Badań) nauczanego w Instytucie Socjologii Uniwersytetu Jagiellońskiego. Na przestrzeni sześciu lat badań terenowych zrealizowano 420 wywiadów uzupełnionych bogatymi danymi metryczkowymi, testami psychospołecznymi, testami stylów komunikowania (zaprojektowanymi przez autora), zdjęciami i innymi materiałami

pozyskanymi od respondentów oraz społeczności gminy Komańcza. Podstawę metodologiczną tego projektu stanowiła teoria ugruntowana i podejście *mixed methods* w analizie CAQDAS. Projekt umożliwił mi nie tylko eksperymentowanie z metodologią i analizą danych, ale pokazał, jak zmienia się świadomość analityczna badacza wraz z ewolucją problematyki badawczej w trakcie kolejnych edycji badań terenowych¹¹. W projektach tych do analizy materiału badawczego wykorzystywałem różne programy CAQDAS (m.in. Atlas.ti, MaxQDA, QDA Miner, Qualrus, HyperResearch, Transana). Użycie programów CAQDAS w analizie danych od początku badań terenowych wpłynęło nie tylko na zmianę sposobu percepcji problemów badawczych w czasie, ale i ukształtowało ramy analityczne ich późniejszej interpretacji socjologicznej¹². Wykorzystując te doświadczenia, chciałbym odnieść się do kilku zagadnień, które są, moim zdaniem, kluczowe dla zrozumienia relacji pomiędzy procesem badań i wspomaganą komputerowo analizą danych jakościowych. Ukazują one bowiem jak duży wpływ wywiera CAQDAS na zmianę sposobu myślenia o metodologii, procesie analizy danych i prowadzeniu terenowych badań jakościowych, a także na rozwój warsztatu badawczego i analitycznego. Logika poniższego wywodu została ujęta w sześciu tezach, które znajdują swoje odbicie w obecnym dyskursie metodologicznym w polu badań jakościowych w socjologii czy szerzej – w naukach społecznych.

¹¹ W kolejnych edycjach badań terenowych udoskonalane były – na bazie prowadzonej regularnie analizy indukcyjnej i abdukcyjnej – techniki, narzędzia (w tym scenariusz wywiadów), procedury badawcze i analityczne.

¹² Efektem tych badań i analiz jest przygotowywana obecnie publikacja książkowa pt. *Jak żyć? Teorie potoczne i strategie życiowe w zbiorowości lokalnej. Studium z socjologii jakościowej*.

CAQDAS w praktyce badań jakościowych

Teza 1: Użyteczność CAQDAS jest w badaniach jakościowych ograniczona przez metodologię procesu badawczego, która implikuje strategię, sposób i procedury analizy danych: ilościowe, jakościowe czy *mixed methods* (metody mieszane)¹³

Niezależnie od tego, czy stosujemy metody manualne, czy wspomaganą komputerowo analizę danych, musimy mieć na uwadze to, że przeprowadzenie analizy jakościowej danych wymaga uważnego, systematycznego i kompleksowego zarządzania dużą ilością danych tekstowych, to jest wywiadów, notatek, zapisów, danych internetowych i tym podobnych. Jeżeli więc warunkiem wstępnym rzetelnej analizy jakościowej jest wydajne i spójne zarządzanie danymi, to zastosowanie nowych technologii, w tym odpowiedniego oprogramowania CAQDAS, staje się naturalne i oczywiste. Rzadko jednak w środowisku badaczy można spotkać się z opinią, że narzędzia do wspomaganą komputerowo analizy danych jakościowych są bazami danych posiadającymi funkcjonalność porządkowania, na przykład kodowania tekstu, wykraczającą poza typowe własności klasycznej bazy danych. Funkcjonalność ta stała się jednak podstawą rozwoju oprogramowania do komputerowej analizy danych jakościowych w latach osiemdziesiątych XX wieku. Napisanie programów CAQDAS przeznaczonych do kodowania i wyszukiwania tekstu oraz

¹³ Metodologia mieszana (*mixed methods*) to podejście, w ramach którego badacz zbiera i analizuje dane, integruje wyniki i wyciąga wnioski, używając jednocześnie metod jakościowych i ilościowych w pojedynczym projekcie lub programie badawczym. W przypadku *mixed methods* metody o charakterze ilościowym i jakościowym powinny być wzajemnie zintegrowane, uzupełniać się, odpowiadać na to samo pytanie badawcze. Koncepcja ta bliska jest pluralizmowi metodologicznemu oraz triangulacji.

opracowanie zaawansowanych funkcji przeszukiwania analizowanych informacji pozwoliło nie tylko na automatyzację rutynowych procedur analitycznych, takich jak zaznaczenie fragmentów tekstu i przypisanie im odpowiednich kodów, ale zapewniło także możliwość łatwego i wielokrotnego wyszukiwania różnych fragmentów analogicznie zakodowanego tekstu bez jego dekontekstualizacji, to jest utraty informacji ułatwiających jego zrozumienie. Współcześnie programy CAQDAS umożliwiają badaczom prowadzenie w formie notatek (*memos*) zapisu wszelkich idei, pomysłów, intuicji pojawiających się w toku realizacji badania empirycznego oraz ich przeglądania w trakcie analizy¹⁴. Programy są również wykorzystywane do wspomaganie procedur analizy danych, ułatwiając badanie własności i relacji w obrębie analizowanych tekstów. Stąd często określa się je mianem programów do budowy teorii, mając na myśli przede wszystkim teorie średniego zasięgu. Nie mam tu na myśli konstruowania, generowania czy odkrywania teorii socjologicznych średniego zasięgu, lecz wykorzystanie oprogramowania CAQDAS w celu rozwijania teorii, to jest przeprowadzania porównań, testowania hipotez czy budowania typologii. Niezależnie od tego, że programy CAQDAS umożliwiają kompleksowe i ustrukturyzowane moderowanie różnymi aspektami analizy danych jakościowych, nie są w stanie wyręczyć badacza z obowiązku czytania materiałów i myślenia. Oprogramowanie CAQDAS jest tylko wsparciem w procesie analizy i budowania teorii opartej na danych jakościowych i nie zastąpi wiedzy, doświadczenia oraz umiejętności badacza. Podobnie jak edytor tekstu sam nie stworzy pliku, tak wspomaganą komputerowo analiza danych

¹⁴ We wcześniejszych wersjach programów CAQDAS ta funkcjonalność była mocno ograniczona.

¹⁰ Zob. Bryda i in. (2010).

jakościowych jedynie ułatwia przeprowadzenie rzetelnej analizy. Sposób analizy danych i interpretacja wyników zawsze leżą po stronie badacza. Stąd kluczem do efektywnego wykorzystywania w analizie danych programów CAQDAS jest przyjęta metodologia badawcza oraz charakter i jakość danych, jakimi badacz dysponuje (Silverman 2007; Lofland i in. 2009; Gibbs 2011). W polu metodologii można wyróżnić kilka dominujących strategii jakościowych, które stosowane są współcześnie w naukach społecznych. Są to między innymi: podejście pragmatyczne, które cechuje eklektyzm metodologiczny służący znalezieniu najlepszej odpowiedzi na problem badawczy, badania etnograficzne/etnometodologia, teoria ugruntowana, fenomenologia, socjologia krytyczna czy badania historyczne i tym podobne¹⁵. Jednakże rozpatrywanie programów CAQDAS tylko z metodologicznego punktu widzenia nie wydaje się właściwe. Proces badawczy i analityczny pozostają bowiem w relacji dialektycznej. Z jednej strony zastosowanie oprogramowania CAQDAS w analizie danych jakościowych stymuluje proces badań terenowych, zaś z drugiej – badania, charakter i jakość zbieranych danych wpływają na procedury i sposób analizy. Proces analizy danych i badania terenowe są ze sobą powiązane (Coffey, Atkinson 1996). Nie powinniśmy jednak zakładać sekwencyjności etapów analizy wobec badań terenowych, lecz ich współzależność (Dey 1993; Di Gregorio, Davidson 2008). Najczęściej algorytm postępowania we wspomaganej komputerowo analizie danych jakościowych składa się z takich etapów, jak:

- planowanie i organizacja bazy danych analitycznych,

- przygotowanie danych do analizy, to jest transkrybowanie, formatowanie, tworzenie procedur analitycznych, czytanie materiału i tym podobne,
- segmentacja i podział danych na mniejsze jednostki analityczne (kodowanie, rekodowanie materiału, pisanie not, przeszukiwanie tekstów lub zakodowanych fragmentów, kategoryzacja, hierarchizacja kodów itp.)¹⁶,
- reorganizacja danych i zakodowanych fragmentów pod kątem problemów badawczych lub kategorii analitycznych, tworzenie diagramów i mapowanie relacji, czyli łączenie danych, kodów, kategorii, linkowanie tekstów, wypowiedzi, zakodowanych fragmentów tekstów i tym podobne,
- tworzenie modeli, typologii i wizualizacja relacji za pomocą tabel kontyngencji, macierzy, ścieżek zależności, zaawansowanych technik prezentacji danych, w tym technik statystycznych (Miles, Huberman 2000; Gibbs 2011; Seale 2013).

Analiza danych jakościowych jest na tyle rozległą dziedziną wiedzy, że trudno znaleźć dla niej wyczerpujący pod względem metodologicznym klucz klasyfikacyjny (Richards 2005). W miarę jak dokonywała się profesjonalizacja badań jakościowych, a co za tym idzie – były one poddawane formalnej kodyfikacji oraz refleksji metodologicznej, w coraz większym

¹⁶ W większości przypadków, ze względu na wymogi czasowe, proces analizy danych jakościowych kończy się na etapie kodowania czy prostych relacji, a często nawet z pominięciem etapów wcześniejszych (Bong 2002).

stopniu następowała ich fragmentaryzacja w obszarze analizy danych (Atkinson, Delamont 2009). Zjawisko to jest efektem coraz większej specjalizacji przedmiotowej i różnorodności tematycznej w obszarze badań jakościowych. Takie zróżnicowanie nie występuje wśród programów do wspomaganej komputerowo analizy danych jakościowych. Nie można jednak zapominać, że to metodologia badawcza, charakter danych czy plan analizy eksploracyjnej powinny określać wybór oprogramowania CAQDAS, a nie odwrotnie. Doskonale pokazują to moje pierwsze doświadczenia z zastosowaniem CAQDAS w badaniach jakościowych, gdzie wybór programu do analizy danych – z braku wiedzy o procesie prowadzenia komputerowej analizy danych i słabej znajomości możliwości CAQDAS – był przede wszystkim podyktowany popularnością programu. Zanim więc dokonamy wyboru programu do komputerowej analizy danych jakościowych, warto skonfrontować ten wybór z osobami, które realizowały już projekty badawcze z wykorzystaniem CAQDAS lub zaznajomić się z publikacjami poświęconymi CAQDAS, ukazującymi ich zastosowanie w konkretnych projektach badawczych oraz dostępne funkcjonalności (Lewins, Silver 2007; Niedbalski 2013). Niestety na gruncie polskim, poza nielicznymi wyjątkami, wciąż brakuje większej liczby rzetelnych opracowań w tym zakresie (Brosz 2012; Niedbalski, Ślęzak 2012).

W doborze oprogramowania do analizy danych jakościowych powinniśmy, po pierwsze, uwzględnić to, jak pozyskujemy dane, a więc przyjętą technikę badawczą: wywiad swobodny, narracyjny, grupowy i tym podobne. Po drugie, wziąć pod uwagę rodzaj danych, jakimi dysponujemy: tekstowe, wizualne czy dane audio. I dalej to, jaki jest charakter tych

danych, przykładowo: narracje i biografie, dane wizualne, dyskurs i dane zwerbalizowane, zbiory materialne, dane archiwalne, dokumenty pisane, listy, strony internetowe, blogi, fora i tym podobne (Silverman 2007; Lofland i in. 2009). W tym sensie metodologia badawcza, jak i przyjęta strategia analityczna (to, co chcemy zrobić z danymi), określa sens wykorzystania danego programu CAQDAS. Zanim więc użyjemy programu CAQDAS w analizie danych jakościowych, musimy odpowiedzieć sobie na kilka pytań: Co jest przedmiotem analizy? Jakie to są dane? Jak są zapisane, utrwalone? Jak będą analizowane? Co ma być efektem analizy danych? Czy dane oprogramowanie nadaje się do celu badania, rodzaju i charakteru posiadanych danych? Czasami zastosowanie oprogramowania CAQDAS może utrudniać proces analityczny przez narzucony, w większości narzędzi, algorytm działania. Szczególnie, gdy pozyskane dane nie do końca nadają się do analizy komputerowej. Zanim badacz skorzysta z wybranego programu CAQDAS w analizie danych jakościowych, powinien wnikliwie rozważyć zasadność jego użycia w odniesieniu do przyjętej metodologii procesu badawczego.

Oprogramowanie w procesie analizy danych jakościowych

Teza 2: Efektywność wykorzystywania CAQDAS w badaniach jakościowych jest odbiciem wiedzy i doświadczenia badacza w analizie danych oraz umiejętności korzystania z różnych funkcjonalności programów

W przeciwieństwie do sytuacji z przełomu lat osiemdziesiątych i dziewięćdziesiątych XX wieku

współcześnie rodzina programów CAQDAS zawiera gamę różnych pakietów narzędziowych dedykowanych szerokiemu spektrum użytkowników¹⁷. W praktyce oprogramowanie CAQDAS można podzielić na trzy grupy. Pierwsza, zdecydowanie największa grupa programów do wspomaganej komputerowo analizy danych jakościowych to licencjonowane, rozwijane od co najmniej dwóch dekad narzędzia wyposażone w wiele złożonych i ciekawych funkcjonalności analitycznych. Druga grupa to bezpłatne narzędzia typu *open source* o mniejszej liczbie dostępnych funkcjonalności, ale zapewniające użytkownikom niezbędne minimum procedur analitycznych (Niedbalski 2013). Trzecia grupa to programy CAQDAS online, rozwijane w ostatnich latach w oparciu o nowe języków programowania. Umożliwiają one analizę danych jakościowych w czasie rzeczywistym z wykorzystaniem metod jakościowych, ilościowych i mieszanych (*mixed methods*), a także dysponują możliwościami konwersji projektów i danych analizowanych w programach komercyjnych. Ich dystrybucja opiera się na zasadach *open source* lub za niewielką opłatą abonamentową (miesięczną/roczną). Programy te wyposażone są praktycznie we wszystkie podstawowe funkcjonalności komercyjnych programów CAQDAS oraz rozwiązania dostępne tylko w formie analiz online. Ich zaletą jest duża elastyczność, prostota i efektywność pracy zespołowej. Poniższa tabela przedstawia zestawienie najpopularniejszych programów z rodziny CAQDAS w porządku chronologicznym ich rozwoju.

¹⁷ Pod szyldem CAQDAS mieści się wiele programów do analizy danych jakościowych, które dysponują podobnymi, podstawowymi funkcjonalnościami. Ich cechą charakterystyczną jest to, co Ann Lewins i Christina Silver określają jako jakościowe podejście do danych jakościowych, a więc: rodzaj i nienumeryczny charakter danych, sposób interpretacji, budowania teorii czy paradygmaty metodologiczno-analityczne (2007: 6–8).

Obecnie wybór programów z rodziny CAQDAS jest szeroki. Każdy wybór oznacza jednak przyjęcie określonej strategii analitycznej, która stoi u podstawy rozwoju danego programu. Wiedza i refleksja użytkowników w tym zakresie są często ograniczone ze względu na słabą znajomość możliwości i funkcjonalności programów w środowisku badaczy jakościowych¹⁸. W Polsce brakuje firm zajmujących się nie tylko dystrybucją oprogramowania CAQDAS, ale i szkoleniami użytkowników. Nie mam tu na myśli szkoleń z zakresu umiejętności korzystania z poszczególnych funkcjonalności danego programu, lecz raczej szkolenia z umiejętności wykorzystywania CAQDAS w badaniach jakościowych¹⁹. Wybór danego programu CAQDAS staje się zazwyczaj indywidualną sprawą każdego badacza. W praktyce, ze względu na koszty komercyjnego oprogramowania, wybór ten ogranicza się zwykle do tych programów, które są popularne lub sprawdzone w środowisku naukowym. Jak pokazują moje własne doświadczenia, świadomość analityczna użycia CAQDAS oraz zdolność wyboru właściwego programu analitycznego rośnie wraz z różnorodnością i częstością realizacji projektów analityczno-badawczych. Wiedza i doświadczenie

¹⁸ Od kilku lat podejmowane są w Polsce działania na rzecz integracji środowiska badaczy jakościowych, np. w formie Transdyscyplinarnych Seminariów Badaczy Jakościowych w naukach społecznych czy działań Sekcji Symbolicznego Interakcjonizmu i Socjologii Jakościowej PTS na gruncie socjologii. Jednakże grono analityków i badaczy zajmujących się profesjonalnie CAQDAS nadal wydaje się być rozproszone. W ramach działań integracyjnych w tym zakresie planowane są w Instytucie Socjologii Uniwersytetu Jagiellońskiego pierwsze warsztaty analityczne w ramach Szkoły CAQDAS z udziałem deweloperów tego oprogramowania.

¹⁹ Szkolenia takie nie są łatwe, ponieważ wymagają wiedzy i sporego doświadczenia badawczego. Choć wydaje się, że trudno oczekiwać od dystrybutorów tego typu szkoleń, to w Niemczech czy Wielkiej Brytanii, gdzie istnieje ścisła współpraca deweloperów i badaczy jakościowych zatrudnionych w jednostkach akademickich, są one regularnie organizowane.

Tabela 1. Rozwój oprogramowania CAQDAS.

ROK WYDANIA	NAZWA PROGRAMU	AKTUALIZACJA PROGRAMU	OBECNA WERSJA	KRAJ POCHODZENIA	FORMA DYSTRYBUCJI	SYSTEM OPERACYJNY / PLATFORMA
1981	N6 (NUD*1ST – do 1997 roku)	Obecnie NVivo	X	Australia	Licencja płatna	Windows
1984	Ethnograph	Cykliczna	6.0.1.0	USA	Licencja płatna	Windows
1987	Aquad	Cykliczna	7.2	Niemcy	GPL	Windows
1989	winMax	Obecnie MAXQDA	X	Niemcy	Licencja płatna	Windows
1991	HyperRESEARCH	Cykliczna	3.5.2	USA	Licencja płatna	Windows, Mac
1993	Atlas.ti	Cykliczna	7	Niemcy	Licencja płatna	Windows, Mac
1995	MAXQDA	Cykliczna	11	Niemcy	Licencja płatna	Windows, Mac
1999	NVivo	Cykliczna	10	Australia	Licencja płatna	Windows, Mac
2000	Kwalian	Nieregularna	5.14	Holandia	Licencja płatna	Windows
2001	Transana	Cykliczna	2.60	USA	GPL + Licencja płatna	Windows, Mac
2002	Qualrus	Nieregularna	X	USA	Licencja płatna	Windows
2002	TAMIS Analyzer	Cykliczna	X	USA	GPL	Mac
2003	Ethnotes	Obecnie Dedoose	X	Kanada	Abonament	web-based
2004	QDAminer	Cykliczna	4.1	Kanada	Licencja płatna	Windows
2006	Xsight (QSR)	Nierozwijany	1.2	Australia	Licencja płatna	Windows
2006	Edeft (Edytor Etnograficzny) WeftQDA	Nierozwijany	X	Polska	GPL	Windows
2006	Coding Analysis Toolkit (CAT)	Nieregularna	X	USA	GPL	web-based
2008	RQDA (dodatek do programu R) Saturate	Cykliczna	X	Chiny	GPL	Windows, Mac, Linux
2009	CATMA	Nieregularna	X	Kanada	Donacja	web-based
2009	Compendium	Nieregularna	1.5.2	Niemcy	GPL	Windows, Mac, Linux, web-based
2010	Olqqa	Nieregularna	X	UK	GPL	Windows, Android
2011	Digital Replay System (DRS)	Cykliczna	X	UK	GPL	Java
2011	Dedoose	Cykliczna	X	USA	Abonament	web-based
2013	LibreQDA	W trakcie rozwoju	X	Hiszpania	GPL	web-based
2013	HyperQualLife (HyperQual)	Nieregularna	X	USA	Abonament	Windows, Mac

* GPL – General Public License; * web-based – Platforma internetowa

Źródło: opracowanie własne. W tabeli uwzględniono najczęściej wykorzystywane przez badaczy programy do wspomaganej komputerowo analizy danych jakościowych w obszarze nauk społecznych.

analityczne wznoszą wraz z procesem kompleksowego korzystania z szeregu funkcjonalności, jakie zapewnia oprogramowanie CAQDAS. Refleksja dotycząca użyteczności danego programu przychodzi dopiero w trakcie procesu analizy danych, interpretacji wyników czy pisania raportów. Wybór oprogramowania powinien więc być dyskutowany już w zespole badawczym na etapie planowania badań terenowych.

Korzystanie z programów CAQDAS podnosi nie tylko świadomość analityczną badacza, ale rozwija i doskonali jego umiejętności informatyczne. Niektóre programy umożliwiają bowiem projektowanie nowych funkcjonalności poprzez pisanie własnych algorytmów czy procedur analitycznych, często niedostępnych z poziomu przeciętnego użytkownika. Rozwijanie tych umiejętności nie tylko jeszcze bardziej przyspiesza rutynowe działania badacza jakościowego, ale przede wszystkim ułatwia mu zapanowanie nad rozproszonym materiałem empirycznym, zmuszając równocześnie do weryfikowania własnego warsztatu, ciągłego uczenia się i doskonalenia posiadanych umiejętności analitycznych. Badacz jakościowy zyskuje dzięki temu właściwy dystans do programu, który staje się jedynie narzędziem wspomagającym proces analizy danych jakościowych, a nie odwrotnie. Praktyka pokazuje bowiem, że bardzo łatwo można wpaść w pułapkę uzależnienia procesu wnioskowania i interpretacji od funkcjonalności danego programu CAQDAS (Brown 2002). Badacz musi być świadomy tego, co robi (lub chce zrobić) i zarazem krytyczny wobec efektów tego, co uzyskuje w procesie analizy danych. Proces generowania czy konstruowania teorii w badaniach jakości-

wych nie odbywa się bowiem w środowisku danego programu CAQDAS, lecz za jego pośrednictwem. Program jest jedynie narzędziem wspomagającym proces rozwijania teorii, a nie narzędziem rozwijającym teorię²⁰.

Wspomniana „pułapka CAQDAS” wydaje się być jedną z tych rzeczy, jakie różnicują badaczy jakościowych na tych, którzy dostrzegają wartość wspomaganej analizy danych i z powodzeniem wykorzystują programy komputerowe w procesie budowania teorii i na tych, którzy bazują raczej na intuicji teoretycznej i doświadczeniu, hołdując manualnym sposobom analizowania danych. W konsekwencji rozwój i profesjonalizacja wspomaganej komputerowo analizy danych jakościowych może pogłębiać podział w środowisku badaczy jakościowych na tych, którzy świadomie decydują się na korzystanie z oprogramowania w analizie i na tych, którzy pozostają wierni metodom manualnym. Wydaje się, że najlepszym sposobem na przełamanie tego podziału jest, z jednej strony, podnoszenie jakości metodologicznej i analitycznej warsztatu badawczego poprzez szkolenia z zakresu wspomaganej komputerowo analizy danych jakościowych, zaś z drugiej – wykorzystywanie wiedzy i doświadczenia tradycyjnie zorientowanych badaczy jako krytycznych. Jedynie zrównoważony dialog tych dwóch stron może przyczynić się do podnoszenia wartości badań jakościowych poprzez wykorzystywanie CAQDAS w socjologii i innych naukach społecznych.

²⁰ Wielu badaczy jakościowych zdaje się zapominać o tym fakcie. Doświadczenie i refleksja metodologiczna podpowiada, że efektywność wykorzystywania programów CAQDAS jest odbiciem znajomości funkcjonalności programów i doświadczenia analitycznego badacza, a nie odwrotnie, jak mogłoby się potocznie wydawać.

Od czego zależy jakość analiz CAQDAS?

Teza 3: Jakość analizy CAQDAS zależy od wolumenu i jakości danych analitycznych

Problemy, które stały się przedmiotem wczesnych debat na gruncie zastosowania wspomaganej komputerowo analizy danych jakościowych w socjologii, znajdują także odzwierciedlenie w dyskursie analitycznym wśród badaczy innych dyscyplin nauk społecznych²¹. Do najważniejszych zalicza się problem rozmiaru wolumenu danych, jakości danych, przygotowania danych do analiz CAQDAS, zarządzania nimi oraz związanej z tymi zagadnieniami rzetelności procedur analitycznych i stylu samej analizy.

Wielkość zbioru, wolumenu danych analitycznych może być w zasadzie nieograniczona. Nie ma znaczenia, czy analizujemy 5, 10 czy 15 wywiadów, czy też odpowiednio więcej zastanych danych tekstowych, na przykład artykułów prasowych, informacji ze stron internetowych, wpisów na portalach społecznościowych czy blogów i tym podobnych. Wielkość zbioru danych może jedynie zwiększać nasycenie kategorii analitycznych czy podnosić rzetelność samej analizy²². Wielu tradycyjnie zorientowanych badaczy jakościowych uznaje, że wykorzystanie CAQDAS w pracy z dużymi zbiorami danych terenowych jest trudne bez utraty

²¹ W obrębie socjologii stanowisko wobec wspomaganej komputerowo analizy danych jakościowych jest podzielone. Sceptycy podnoszą podobne argumenty, jakie pojawiały się w dyskusjach w wśród anglosaskich socjologów jakościowych na przełomie lat osiemdziesiątych i dziewięćdziesiątych XX wieku.

²² Wielkość zbioru danych w programach CAQDAS jest zazwyczaj konsekwencją celu i metodologii badania jakościowego, oczekiwań badacza czy wymogiem ze strony zleceniodawcy badań.

szczegółowości informacji zawartych w danych, a co za tym idzie – może ograniczać rzetelność analizy. Doświadczenie uczy, że praca na dużych zbiorach danych jakościowych prowadzi do tego, iż badacz w analizie CAQDAS zaczyna koncentrować się bardziej na szerokości niż na głębokości danych (Seidel 1991). Jak bowiem poradzić sobie percepcyjnie i organizacyjnie z sytuacją, gdy zrealizowaliśmy 150 wywiadów, po około 1,5 godziny każdy, co szacunkowo daje 3000–4000 stron transkrypcji, a przy zastosowaniu CAQDAS w procesie kodowania materiału wystarczająco dużą liczbę zakodowanych fragmentów danych (oczywiście w zależności od tego, co przyjmiemy jako jednostkę kodowania). Liczba zakodowanych fragmentów rośnie proporcjonalnie do wielkości jednostki kodowania. Jeśli kodujemy dane według reguły *incident-by-incident*, jest ona zdecydowanie mniejsza, jeśli zaś według reguły *line-by-line* – rośnie lawinowo. Przykładowo, we wspomnianym wcześniej projekcie o atrybucjach społecznych błędną decyzją okazało przyjęcie reguły kodowania *line-by-line* (Charmaz 2009; Gibbs 2011). Po wstępnym zakodowaniu ponad 200 wywiadów otrzymano około 25 000 zakodowanych fragmentów wywiadów. W przypadku większych wolumenów danych wstępna analiza danych jakościowych powinna być raczej zredukowana do eksploracji tematycznej – ze względu na trudności wynikające z ograniczeń percepcyjnych tak obszernych treści. Warto w takiej sytuacji rozważyć kodowanie większych fragmentów tekstu, o ile jest to zgodne z celami projektu badawczego.

Szerokość danych, to jest odpowiednia liczba zbioru danych, w przeciwieństwie do głębokości daje

badaczowi wrażenie większej wiarygodności procesu analizy. Podobnie jak reprezentatywność próby w badaniach ilościowych pozwala na traktowanie wyników analizy jako „pewników”, a jej samej nadaje pozory rzetelności. Dla wielu badaczy praca na dużych zbiorach danych jest jednak zaprzeczeniem istoty metodologii badań jakościowych, utratą tego, co stanowi siłę podejścia jakościowego, to jest szczegółowego wglądu w sytuację społeczną, przedmiot badania. Jak wskazują Fielding i Lee (1998), w projektach jakościowych przeciętna wielkość zbioru danych i próby badawczej nie rośnie lawinowo. Ich zdaniem to nie dostępność oprogramowania powoduje, że badacze realizują duże projekty, lecz potrzeba realizacji dużych projektów zmusza ich do szukania odpowiedniego oprogramowania wspomagającego proces zarządzania danymi i analizą. Innymi słowy, jeśli chcemy w małych badaniach jakościowych wykorzystać oprogramowanie CAQDAS, powinniśmy się zastanowić nad sensownością takiego rozwiązania. Seale (2001) kwestionuje zasadność użycia CAQDAS w odniesieniu do mało licznych zbiorów danych. W praktyce badań jakościowych, jak wskazuje doświadczenie metodologiczne i analityczne, badacz powinien raczej szukać równowagi między wielkością wolumenu danych, ich szerokością a głębokością analizy. W przeciwnym razie albo nie będzie zdolny do zapanowania nad posiadanymi danymi, albo zbyt dużą szczegółowość w analizie nie pozwoli mu oderwać się od pierwotnego kontekstu znaczeniowego danych. Rozsądne użycie CAQDAS w badaniach jakościowych jest więc wypadkową pomiędzy umiejętnościami organizacji danych, zarządzania materiałem empirycznym i stopniem głębokości planowanej przez badacza analizy.

Zarządzanie wolumenem danych terenowych: transkrypcjami z wywiadów, notatkami, nagraniami audio, wideo, zdjęciami i tak dalej, a w konsekwencji procesem ich analizy, nie byłoby możliwe bez użycia specjalistycznego oprogramowania komputerowego. Wsparcie CAQDAS pozwala na szybką techniczno-administracyjną obróbkę danych, ich przygotowanie do analizy (podział na zestawy, paragrafy, sekcje itp.) czy kodowanie. Kształtuje nowy sposób pracy z danymi jakościowymi, który byłby trudny do realizacji za pomocą metod manualnych (Tesch 1990; 1991). Dzięki CAQDAS badacz pełniej trzyma się planu analizy, może eksperymentować z danymi, kodować je i rekodować w dowolny sposób, zgodnie z przyjętą logiką analizy, testować pojawiające się w trakcie pracy hipotezy, idee i pomysły analityczne. Zdaniem Fieldinga i Lee (1998) użycie CAQDAS w badaniach jakościowych nie tylko redukuje pracę papierową, ale pomaga skupić uwagę badacza na ukrytych wymiarach analitycznych rozproszonych danych. Ponadto jeśli dane są wewnętrznie uporządkowane, sensownie zorganizowane i zarządzane, mogą być użyte w nowych celach badawczych czy analitycznych (*secondary data analysis*). Doświadczenie pokazuje, że praca z dużymi zbiorami danych jakościowych zdecydowanie rozwija szereg umiejętności analitycznych. Jednakże trudno pominąć fakt, że praca nad kilkoma wywiadami pogłębia umiejętności rozumienia tła społeczno-kulturowego analizowanych danych, osadza je w szerszym kontekście znaczeniowym. Praca z dużymi zbiorami danych rozszerza perspektywę analityczną, ale praca z małymi zbiorami uczy rzetelności warsztatowej i lepszego rozumienia procesu wspomaganego komputerowo analizy danych jakościowych.

Niezwykle istotną cechą wykorzystywania CAQDAS jest możliwość replikacji, to jest odtworzenia procesu badań jakościowych i analizy danych, co sprzyja ich większej transparentności (Conrad, Reinharz 1984b). W tradycyjnym modelu analizy danych jakościowych zadania te są utrudnione i czasochłonne, a warsztat badacza pozostaje zazwyczaj ukryty. Korzystanie z programów CAQDAS pozwala na pełną automatyzację rutynowych działań, oszczędność czasu i efektywność pracy analitycznej (Tesch 1990; 1991). Badacz może łatwiej skupiać uwagę na analizowanych danych i znajdować w nich to, co interesujące i istotne. Przeglądanie oraz przeszukiwanie licznych danych, kodów, notatek czy zakodowanych fragmentów staje się procesem wielokrotnie powtarzającym i ciągłym. Doświadczenie projektowe pokazuje, że zmusza to badacza do systematyczności i retrospektywnego podejścia w procesie analizy, stałego weryfikowania własnych interpretacji poprzez ich ugruntowanie analityczne, zakorzenianie w danych. CAQDAS pozwala również na względnie łatwe identyfikowanie przypadków odstających czy wybieranie do analizy mniejszych jednostek danych ukrytych w masie informacji.

Niezależnie od tego, czy nasz zbiór danych jest duży czy nie, musimy mieć na uwadze problem rzetelności i jakości wspomaganego komputerowo analizy danych. Jakość tej analizy jest oczywiście konsekwencją jakości uzyskanego materiału empirycznego, ale jest również związana z dokumentowaniem procedur kodowania, procesu analizy i wnioskowania. To czasochłonne, wymagające wysiłku i pracy zadanie staje się w miarę nabierania doświadczenia w pracy z programami

CAQDAS działaniem rutynowym. Dzięki temu metodologia i procedura badawcza, a w szczególności etapy i sposób analizy danych, stają się przejrzyste dla innych badaczy jakościowych.

Styl komputerowej analizy danych

Teza 4: Styl analizy CAQDAS jest kombinacją wiedzy metodologicznej, wyobraźni socjologicznej oraz doświadczenia badawczego i analitycznego

W badaniach jakościowych, niezależnie od tego, czy używamy programów CAQDAS czy też tradycyjnej metodologii, sposób podejścia do danych, patrzenia na nie, dostrzegania i eksponowania zawartych w nich relacji można nazwać stylem analizy danych. Styl analizy danych jakościowych jest czymś, co każdy badacz rozwija zarówno w zaciszu własnego gabinetu, jak i w terenie. Posiadanie własnego stylu daje poczucie pewności siebie, możliwość panowania nad materiałem badawczym, a dla innych jest oznaką profesjonalizmu badacza. Styl analizy jest raczej cechą osobistą, a nie grupową²³. Doświadczenie projektowe pokazuje, że jest on kombinacją posiadanej wiedzy metodologicznej, praktyki w realizacji projektów badawczych oraz umiejętności pracy z danymi, a co za tym idzie – przekładania wiedzy teoretycznej i metodologicznej na procedury i konkretne działania analityczne.

W tym sensie styl analizy CAQDAS (lub krócej – styl CAQDAS) można określić jako relację między konceptualizacją badawczą wyrażoną w postaci

²³ W przypadku grupy badaczy można raczej mówić o szkole niż stylu analizy danych jakościowych.

wiedzy teoretycznej, dyskursywnej, wiedzy typu „co” (*know*) a operacjonalizacją, to jest wiedzą proceduralną, wiedzą praktyczną typu „jak” (*how*), za pośrednictwem przez dany program do analizy danych jakościowych. Styl CAQDAS to sposób pracy z danymi, wnioskowania i interpretacji, w którym kluczową rolę odgrywa metodologia i umiejętność radzenia sobie z programami do wspomaganej komputerowo analizy danych jakościowych.

Istotną cechą stylu wspomaganej komputerowo analizy danych jakościowych jest umiejętność eksperymentowania z danymi, unikanie nadmiernego schematyzmu w procesie analizy, ciągłe poszukiwanie i weryfikowanie wzorców oraz relacji tkwiących w danych. Styl CAQDAS jest więc czymś unikalnym, wizytówką badacza jakościowego, wskaźnikiem jego profesjonalizmu i umiejętności radzenia sobie z danymi. To wszystko zmusza badacza do nieustannej refleksji, bycia uważnym w poszukiwaniu relacji tkwiących w danych i otwartym na pojawiające się nowe konfiguracje danych. Rola CAQDAS w rozwijaniu stylu analizy danych jakościowych jest dwojaka. Z jednej strony etapy procesu badawczego oraz postępująca proceduralizacja i profesjonalizacja analizy danych jakościowych kształtują osobowość i tożsamość badacza oraz jego sposób pracy z danymi. Z drugiej strony użycie komputerów w analizie może prowadzić do kreowania dystansu między badaczem a danymi. Wielu badaczy jakościowych stosujących oprogramowanie w analizie danych uważa jednak, że dzięki CAQDAS czują się zdecydowanie bliżej analizowanych danych (Weitzman, Miles 1995). Koncepcję „bliskości danych” i jej sens metodologiczny kwestionują Fielding i Lee (1998). Wskazują,

że użycie metod manualnych nie gwarantuje bliskości względem danych, gdyż ta jest najczęściej funkcją wielkości wolumenu danych. Ich zdaniem bliskość danych oznacza aktualne, fizyczne obcowanie z danymi, bezpośredni dostęp do danych²⁴. Z kolei Ragin i Becker (1989) argumentują, że CAQDAS zapewnia bliskość i interaktywny styl analizy danych jakościowych. Argument wyjątkowości i niepowtarzalności bezpośredniego kontaktu z danymi, ich bliskości oraz kontekstu znaczeniowego, cechujący analizy manualne, wydaje się tracić na znaczeniu, gdyż dzięki zastosowaniu CAQDAS możemy prześledzić metodologię i proces analizy, a tym samym doskonalić jej rzetelność. Jeżeli trafność jest cechą badań jakościowych, to dzięki CAQDAS możemy rozwijać rzetelność analizy, szczególnie przy analizie dużych wolumenów danych. Trafność i rzetelność harmonizują nie tylko styl analizy danych jakościowych, ale czynią procedury i proces analizy przejrzystymi dla szerszego grona badaczy²⁵. Jak pokazuje doświadcze-

²⁴ Stanowisko badaczy jakościowych jest w tym względzie podzielone. Nie ma zgody co do tego, czy CAQDAS ułatwia, czy utrudnia proces analizy. Czasami korzystanie w analizie danych z CAQDAS może być trudne, gdy oprogramowanie jest niedostosowane do rodzaju danych. Fielding i Lee (1998) wskazują na to, że wszelkie przekształcanie danych wymaga dobrej znajomości ich kontekstu znaczeniowego, co było utrudnione w pierwszych programach CAQDAS. Natomiast Weitzman i Miles (1995) sugerują, że dystans wobec danych można zredukować przez wprowadzenie odpowiedniego interfejsu programu analitycznego.

²⁵ Doświadczenia z CAQDAS wskazują na szereg problemów, które utrudniają użycie oprogramowania w analizie danych jakościowych. Są to zarówno problemy metodologiczne, jak i techniczne. Przykładowo, istnieje szereg poprawnych ścieżek metodologicznie wspomaganej komputerowo analizy danych. Spotkania konferencyjne i seminaryjne w obszarze badań jakościowych pokazują, że badacze korzystający z oprogramowania CAQDAS nie mają do końca wyobraźni, co mogą, poza kodowaniem czy przeszukiwaniem tekstów, zrobić z posiadanymi danymi. Trudno jest bowiem przełożyć koncepcję analizy na konkretną procedurę czy funkcjonalność programu CAQDAS. Dodatkowo mogą dochodzić do tego problemy z zarządzaniem rozproszonymi źródłami danych czy wielkością wolumenu danych.

nie projektowe, CAQDAS uczy i wymaga precyzji w procesie analizy, wyobraźni analitycznej, czego najlepszym wskaźnikiem są jego efekty.

W praktyce styl komputerowej analizy danych jakościowych jest powiązany z procesem ich kodowania. Podobnie jak w przypadku kodowania, styl CAQDAS można sprowadzić do trzech strategii postępowania badacza z materiałem empirycznym: indukcyjnej, dedukcyjnej i mieszanej. Strategia indukcyjna w analizie danych jakościowych cechuje podejście od strony materiału empirycznego, operacjonalizacji terenowej danych do ich konceptualizacji w analizie. Jej celem jest generowanie nasyconych treścią kategorii analitycznych poprzez redukcję kontekstu znaczeniowego analizowanych danych do niezbędnego minimum. Strategia dedukcyjna odnosi się operacjonalizacji kategorii analitycznych na bazie klucza kategoryzacyjnego, powstałego na przykład w oparciu o scenariusz wywiadu. Jej celem jest egzemplifikacja empiryczna kategorii analitycznych, to jest poszukiwanie wskaźników w danych terenowych, które zgodnie z logiką klucza kategoryzacyjnego odpowiadają tym kategoriom. Strategia mieszana stara się łączyć oba podejścia, przyjmując jednak za punkt wyjścia ścieżkę indukcyjną.

Kodowanie – rdzeń analizy CAQDAS

Teza 5: Styl CAQDAS jest odzwierciedleniem jakości i trafności kodowania

Niezależnie od tego, jaki przyjmujemy rodzaj strategii analitycznej, musimy odpowiedzieć na kilka podstawowych pytań, między innymi: Ja-

kiego rodzaju dane analizujemy? Jak możemy je scharakteryzować? Jakie są nasze cele analityczne? Dlaczego wybraliśmy te dane do analizy? Kto i czego chce się dowiedzieć z tej analizy? Każda odpowiedź na te przykładowe pytania nie tylko kształtuje naszą strategię analityczną, ale ma odzwierciedlenie w wyborze zarówno jednostki, jak i sposobu kodowania danych jakościowych. Przy czym jednostką kodowania może być słowo, zdanie, paragraf, wypowiedź, obraz czy jego fragment i tym podobne. Procedura kodowania stanowi więc rdzeń wspomaganej komputerowo analizy danych (Bong 2002). Kodowanie i analiza danych powinny być rozwijane już w trakcie realizacji badań terenowych. Takie podejście do procesu kodowania danych narzuca najbardziej wpływowy paradygmat w analizie CAQDAS – teoria ugruntowana. Od samego bowiem początku – zgodnie ze strategią indukcyjną stojącą u podstaw logiki analizy w teorii ugruntowanej – zbieranie danych, ich kodowanie i sama analiza to działania współzależne, rozwijane równocześnie²⁶. Kodowanie jest rdzeniem każdej analizy jakościowej, ale jego efektywność zależy od wielkości wolumenu danych, jakimi dysponujemy. W praktyce analizy CAQDAS im większy jest wolumen danych jakościowych, tym bardziej uzasadnione jest zastosowanie procedur kodowania dedukcyjnego. Natomiast kodowanie indukcyjne lepiej sprawdza się na mniejszych zbiorach danych. Jest uzasadnione na etapie konstruowania

²⁶ Istnieją też inne podejścia odnoszące się do współzależności tych trzech elementów. W praktyce trudno bowiem wyobrazić sobie tę współzależność, gdy przedmiotem analizy są dane zastane, tak jak chociażby w analizie treści czy zawartości przekazów masowych. Niezależnie od tego diada kodowanie – analiza jest podstawą wspomaganej komputerowo analizy danych jakościowych.

i testowania klucza kategoryzacyjnego lub struktury kodów analitycznych, które będą później stanowić wzorzec dedukcyjnego kodowania danych. Przed przystąpieniem do kodowania danych musimy sprawdzić, jaki jest wolumen, rodzaj i jakość danych, którymi dysponujemy, określić jednostkę kodowania i analizy, a w konsekwencji – wybrać sposób kodowania danych zgodnie z wspomnianym poprzednio podejściem indukcyjnym lub dedukcyjnym. Podobnie jak proces wspomaganą komputerowo analizy danych jakościowych wymaga od badacza czasu i ciągłej refleksji analitycznej, tak i kodowanie wymaga cierpliwości, systematyczności i bycia otwartym na materiał badawczy. Na każdym etapie analizy rodzaj i sposób kodowania w programach CAQDAS są w pełni kontrolowane przez badacza, lecz ich efektywność zależy od rodzaju posiadanego materiału. Inna jest specyfika kodowania danych tekstowych: wywiadów narracyjnych, biograficznych, swobodnych czy artykułów prasowych, a inna danych wizualnych: obrazów, zdjęć czy filmów. Doświadczenie projektowe uczy, że sposób kodowania powinien być wyraźnie określony przez cel badań. Jego brak utrudnia myślenie o materiale badawczym i rodzi chaos analityczny, co w konsekwencji wymaga powtórnego przekodowania już raz zakodowanych danych. Z problemem tym miałem do czynienia we wstępnej fazie realizacji projektu o atrybucjach społecznych. Po kodowaniu otwartym techniką *line-by-line* okazało się, że przyjęta przeze mnie jednostka kodowania dla wywiadu swobodnego jest nietrafna, ponieważ ten rodzaj kodowania powoduje utratę kontekstu znaczeniowego. Celem badania terenowego była eksploracja tematyki, a więc sensowniejsza oka-

zała się w tej sytuacji technik kodowania *incident-by-incident*. Doświadczenie to nauczyło mnie, że sposób kodowania jest ściśle związany nie tylko z celem badań, ale i z przyjętą metodologią (techniką) badania terenowego.

Kodowanie w programach CAQDAS jest procesem iteracyjnym, w którym struktura kodów jest dynamiczna, ulega ciągłym przekształceniom wraz z zagłębianiem się badacza w konteksty znaczeniowe i strukturę semantyczną danych jakościowych. Jest działaniem, które daje dużą dozę satysfakcji, lecz zabiera długie godziny spędzone przed komputerem²⁷. Dokonuje się indywidualnie lub zespołowo, w zaciszu gabinetów naukowych, tam gdzie liczy się wiedza, intuicja, doświadczenie badawcze i umiejętności analityczne. Nie ulega wątpliwości, że kodowanie jest fundamentem wspomaganą komputerowo analizy danych jakościowych. Obecnie, wraz z rozwojem nowych funkcjonalności w programach CAQDAS, większy nacisk jest położony na procedurę walidacji danych i rzetelności kodowania, tworzenie powiązań i poszukiwanie relacji logicznych pomiędzy kodami, fragmentami zakodowanych danych czy dokumentami (hiperlinkowanie), a także ich wizualizację czy geolokalizację. Procedury te stanowią ogromne wsparcie dla badacza-analityka w procesie generowania czy konstruowania teorii. Pozwalają wydobyć na światło dzienne ukryte wymiary analizowanych zjawisk czy procesów społecznych.

²⁷ Kodowanie jest bardzo mocno zakorzenione w praktyce analizy CAQDAS, co powoduje, że mało jest przykładów wychodzenia poza proces kodowania, względnie klasyfikacji danych. Stąd efektem CAQDAS są często uogólnienia empiryczne, z rzadka pretendujące do rzetelnych analiz teoretycznych.

CAQDAS. Od eksploracji danych do eksplanacji teoretycznej

Teza 6: Użycie CAQDAS w procesie badawczym ułatwia procesy interpretacji i rozwijania teorii, podnosi rzetelność analizy danych, a także sprzyja rozwijaniu umiejętności analitycznych i informatycznych, bez których trudno wyobrazić sobie współczesne, jakościowo zorientowane badania socjologiczne

Niezwykle trudno nauczyć wspomaganą komputerowo analizy danych jakościowych teoretycznie, z książek. Jeśli badacz chce poznać specyfikę i możliwości CAQDAS, musi uzbroić się w cierpliwość i rozpocząć od podstaw naukę obsługi oprogramowania oraz procedur analitycznych. Moje doświadczenia w tym zakresie pokazują, że świadomość analityczna i umiejętność wykorzystywania CAQDAS w badaniach jakościowych rosną szybciej, gdy używamy oprogramowania do analizy danych jakościowych, realizując własne, małe projekty badawcze²⁸. CAQDAS wymaga od badacza ciągłego doskonalenia umiejętności analitycznych poprzez eksperymentowanie z funkcjonalnościami danego programu. Oprogramowanie nie dyktuje jednak procedur czy etapów analizy. Te zależą od samego badacza i świadomości tego, co chce zrobić oraz tego, jak może wykorzystać dany program. Innymi słowy, uczenie się pracy z CAQDAS wymaga zmiany stylu myślenia o procesie badań i analiz jakościowych. Przystępując do analizy danych z użyciem CAQDAS, badacz musi

²⁸ Szkolenia z CAQDAS organizowane przez dystrybutorów oprogramowania czy badaczy jakościowych nastawione są raczej na uczenie umiejętności operowania funkcjonalnościami danego programu.

również zdecydować się na przyjęcie określonego paradygmatu analitycznego. Najczęściej jest to wybór pomiędzy tradycyjnym realizmem metodologicznym, zakładającym odrzucenie subiektywnej rzeczywistości lub jego umiarkowaną wersją – społecznym konstrukcjonizmem, nazywanym także konstruktywizmem, zgodnie z którym rzeczywistość nie jest oderwana od znaczeń i kontekstów kulturowych²⁹.

Wielu badaczy jakościowych wykorzystuje programy CAQDAS przede wszystkim do wielowymiarowej eksploracji danych jakościowych, integracji i zarządzania bazą danych w celu stworzenia ram dalszej analizy. Proces ten służy zrozumieniu danych oraz znajdowaniu struktury relacji pomiędzy kategoriami analitycznymi. Eksploracja jest procesem iteracyjnym, służącym weryfikacji struktury kodów i ścieżki analizy. Wymaga ona ciągłej refleksji nad procesem analizy – weryfikacji i zmian w podejściu, elastyczności i otwartości na nowe znaczenia. Programy CAQDAS są wyposażone w szereg funkcjonalności (m.in. wyszukiwania i przeszukiwania treści czy zakodowanych fragmentów danych), które ułatwiają eksplorację danych jakościowych.

Rozwój dodatkowych funkcjonalności, chociażby takich jak mapy kognitywne, zestawienia tabelaryczne czy reguły współwystępowania, dostępnych głównie w komercyjnych programach CAQDAS, umożliwia również rozwijanie procesu

²⁹ Zgodnie ze stanowiskiem realizmu metodologicznego, fakty społeczne – rzeczywistość empiryczna – istnieją niezależnie od umysłu badacza i są dostępne poznaniu naukowemu, podczas gdy przedstawiciele społecznego konstrukcjonizmu uznają, że fakty – rzeczywistość empiryczna – są społecznie konstruowane.

eksplanacji i teoretyzowania (Miles, Huberman 2000; Lewins, Silver 2007). Eksplanacja jest tym etapem analizy danych, który wymaga od badacza już określonej wiedzy, znajomości technik analitycznych, a także wspomnianej poprzednio wyobraźni socjologicznej. To na tym etapie analizy wykorzystywane są dodatkowe informacje, na przykład dane metryczkowe respondentów z wywiadów, pojawiają się modele konceptualne, mapy poznawcze, klasyfikacje, zestawienia i diagramy analiz przekrojowych, a także wszelkie zaawansowane formy wizualizacji. A gdy kodowanie materiału znajduje przełożenie na język zmiennych, nominalnych czy porządkowych, otwierają się możliwości wykorzystywania w analizie CAQDAS technik znanych z analiz statystycznych, na przykład analiz skupień czy analizy korespondencji. Coraz częściej w programach tych implementowane są metody *mixed methods*, w których łączy się podejście jakościowe i ilościowe. W podejściu tym łączy się, przy zachowaniu specyfiki metodologii zbierania danych właśnie na etapie eksplanacji i interpretacji, zarówno różne typy danych, jak i techniki ich analizy. Dzięki temu uzyskujemy możliwość poszerzenia i pogłębienia analizy jakościowej. Ten swoisty pragmatyzm metodologiczny oparty na funkcjonalnościach CAQDAS pozwala na kompleksową analizę danych i ugruntowane metodologicznie konstruowanie jakościowych teorii średniego zasięgu.

Przechodzenie przez kolejne etapy analizy CAQDAS, od eksploracji do eksplanacji, wymaga świadomości całego procesu analizy i dostrzegania jego związku z całością zmian procesu badawczego. Wspomagana komputerowo analiza

danych jakościowych nie jest tylko kolejnym elementem czy etapem procesu zbierania danych. Jest niezależnym, lecz stale współtowarzyszącym mu procesem organizacji danych i przechodzenia na coraz wyższe poziomy rozumienia ich kontekstu znaczeniowego. Przejścia te są płynne i często trudne do weryfikacji, dlatego istotnym elementem CAQDAS, przynajmniej w fazie kodowania i eksploracji danych, może być praca zespołowa. Zespołowość jest szczególnie pożądana na etapach wstępnych analizy danych jakościowych, choć w praktyce rodzi ona wiele problemów interpretacyjnych, szczególnie ze względu na występujące wśród badaczy różnice w rozumieniu kontekstów znaczeniowych. Doświadczenia w pracy z zespołami studenckimi pokazują, że różnice te są najczęściej pochodną cech psychologicznych badacza, na przykład wrażliwości, empatii, cech interpersonalnych lub cech społeczno-kulturowych, takich jak schematy poznawcze czy kapitał kulturowy, edukacyjny, lingwistyczny. Przykładowo, istotnym elementem w procesie kodowania wywiadów jest kapitał lingwistyczny osób kodujących dane jakościowe, który może ułatwić lub utrudnić dalsze etapy analizy. Jego brak powoduje, że kodowanie danych jest, używając określenia z języka potocznego, „płaskie”, co zwykle oznacza, że badacz nie do końca rozumie kontekst znaczeniowy wypowiedzi respondenta i nietrafnie (zbyt wąsko lub zbyt szeroko) koduje ich sens. Podobnie doświadczenie projektowe uczy, że gdy dysponujemy dużym wolumenem danych, sensowniejszym i efektywniejszym rozwiązaniem, szczególnie na etapie kodowania otwartego czy eksploracji danych, jest praca zespołowa. Członkowie zespołu badawczego dysponują zazwyczaj zróżnicowanymi zasobami,

doświadczeniem i umiejętnościami analitycznymi, ale interaktywność, która jest unikalną cechą zespołów badawczych, sprzyja rozwijaniu i doskonaleniu umiejętności pracy z danymi i programami CAQDAS.

Niezależnie od tego, czy prowadzimy analizę danych indywidualnie czy zespołowo, praca z CAQDAS wymaga od badacza wiedzy, praktyki i określonych umiejętności, zdolności zarówno do analizy, jak i syntezy. CAQDAS jest tylko narzędziem w procesie analizy, wsparciem dla badacza, ale nie zwalnia go z myślenia i interpretowania sensu analizowanych danych. Można więc powiedzieć, że eksploracja danych jest rzemiosłem, w którym odbija się warsztat analityczny badacza jakościowego, a eksplanacja sztuką interpretacji, która ukazuje jego poziom wyobraźni socjologicznej oraz umiejętność socjologicznego teoretyzowania.

Podsumowanie

Metodyka pracy z programami CAQDAS w procesie badań jakościowych uczy dyscypliny i podnosi świadomość analityczną badacza, zwracając jego uwagę na nowe wymiary i konfiguracje danych. Zmusza badacza do trzymania się określonych procedur analitycznych, do weryfikowania stawianych tez, wielowymiarowego myślenia o materiale badawczym, a w konsekwencji do wypracowywania własnej ścieżki analitycznej, to jest sposobu pracy z danymi. Z jednej strony uczy myślenia analitycznego o materiale badawczym, z drugiej – na przykład przez stosowanie metody ciągłego porównywania – myślenia syntetycznego pozwalającego na oderwanie się od/dystans do danych

i przechodzenie w procesie kodowania od kategoryzacji do konceptualizacji. Jak pokazuje doświadczenie dydaktyczne i empiryczne, bycie dobrym analitykiem w dziedzinie wspomaganego komputerowo analizy danych jakościowych to sztuka³⁰. CAQDAS wymaga bowiem od badacza jakościowego elastyczności w myśleniu, wyobraźni socjologicznej i analitycznej, wysokiego poziomu wrażliwości językowej, wejścia w interakcję z danymi, dostrzegania i rozumienia wielowymiarowości relacji między nimi. Praca z programami CAQDAS wymaga cierpliwości i systematyczności, szczególnie wtedy, gdy uświadomimy sobie, że w praktyce co najmniej 80 procent czasu pracy analitycznej zajmuje przygotowanie danych i kodowanie, a reszta to rzeczywista analiza danych (choć zgodnie z logiką teorii ugruntowanej, kodowanie to także element analizy). Korzystanie z CAQDAS kształtuje sposób patrzenia na dane, proces analizy i jego etapy, ale przede wszystkim podnosi wartość samej analizy, nadając jej status wiarygodności. Choć z drugiej strony rzetelna analiza CAQDAS wymaga od badacza bardzo dobrej znajomości programów, czasu i doświadczenia w ich używaniu. Dlatego w praktyce niewielu badaczy dochodzi w procesie analizy CAQDAS do konstruowania modeli teoretycznych, poprzestając, ze względu na jej czasochłonność, na wykorzystywaniu programów do kodowania lub tworzenia prostych typologii. Pojawia się więc pytanie, kiedy używać programów CAQDAS w procesie badań jakościowych. Doświadczenie analityczne podpowiada, że używanie oprogramowania

³⁰ Doświadczenie dydaktyczne pokazuje, że bycie dobrym badaczem terenowym nie zawsze idzie w parze z umiejętnościami analizy CAQDAS. Niezależnie od tego, podobnie jak badania jakościowe, praca z programami CAQDAS doskonali i poszerza zakres wyobraźni socjologicznej.

CAQDAS jest sensowne przede wszystkim wtedy, gdy materiał badawczy przerasta możliwości percepcyjne badacza jakościowego, gdy trudno jest tak uporządkować dane, by nadać im sens analityczny. Programy CAQDAS pełnią wtedy najczęściej funkcję bazy, która służy organizacji i zarządzaniu danymi jakościowymi. Z drugiej strony trudno jednoznacznie wskazać, czy 15, 20 wywiadów to wystarczający zbiór danych, by korzystać z programów do wspomaganego komputerowo analizy danych jakościowych. Jednakże stosowanie oprogramowania CAQDAS na mniejszych zbiorach danych pozwala badaczowi na wypracowanie własnego stylu analitycznego, doskonalenie warsztatu czy testowanie reguł metodologicznych pracy z danymi jakościowymi, znanych często tylko z teorii analizy danych³¹. Moje własne doświadczenia pokazują, że programy CAQDAS rozwijają umiejętność analizy danych jakościowych w dużo większym stopniu niż metody manualne. Jeśli jednak chcemy rzetelnie korzystać z oprogramowania CAQDAS w badaniach jakościowych, powinniśmy, zanim rozpoczniemy komputerową analizę danych, przeprowadzić analizę w sposób tradycyjny. Takie podejście do analizy danych pozwala bowiem na zrozumienie procesu analitycznego, poznanie procedur i uczy zasad pracy z danymi jakościowymi. Dzięki temu badacz staje się bardziej świadomy etapów i przebiegu analizy, rozpoczynając pracę z programami CAQDAS.

Podsumowując powyższe rozważania, chciałbym wskazać kilka zalet i ograniczeń korzystania z pro-

³¹ W praktyce, w zależności od wielkości zbioru danych, możemy wyróżnić dwa podejścia do analizy CAQDAS. Albo analizujemy mało liczne zbiory danych w formie pogłębionego *case study*, albo przedmiotem analizy są duże zbiory wymagające od badacza doświadczenia i określonych umiejętności analitycznych.

gramów CAQDAS w praktyce badań jakościowych. Podstawową zaletą jest usystematyzowanie relacji pomiędzy procesem badania terenowego i analizy danych. Relacja ta od początku realizacji badań jest współzależna. CAQDAS podnosi profesjonalizm analizy danych. Na każdym jej etapie badacz dysponuje jasnymi procedurami, wie kiedy, co i jak robić. Z drugiej strony interpretacja danych w metodologii jakościowej przestaje być tylko konstrukcją myślową, a staje się konsekwencją rzetelnej pracy terenowej, stylu i warsztatu analitycznego. CAQDAS ułatwia również kompleksową integrację różnych rodzajów danych jakościowych oraz weryfikację ich jakości, a więc i jakości pracy terenowej. Dzięki dużej elastyczności programów istnieje możliwość ciągłej, dowolnej reorganizacji i zarządzania danymi, strukturą kodów, kategoriami analitycznymi i procesem analizy. CAQDAS to także sposób na dokumentowanie procesu badawczego i analitycznego, dzięki czemu istnieje możliwość odtworzenia ścieżki analizy jakościowej, powtórzenia jej etapów przez innego badacza, co w konsekwencji podnosi rzetelność wyników. Systematyzacja procedur i automatyzacja rutynowych operacji manualnych zapewniają kontrolę nad kolejnymi etapami analizy. Proces badania jakościowego staje się otwarty dla innych osób. Praca z programami CAQDAS nie tylko rozwija i doskonali warsztat badawczy oraz analityczny, ale i uczy niezbędnego w analizie danych dystansu do materiału badawczego. Dostępność różnych funkcjonalności CAQDAS ułatwia nowe spojrzenie na problemy badawcze. Badacz uzyskuje szerszą perspektywę analityczną i teoretyczną, nie jest całkowicie „zanurzony” w danych jakościowych. Korzystanie z szeregu funkcjonalności wymaga jednak dobrej znajomości programów

CAQDAS, a tego nie można nauczyć się teoretycznie. Tak, jak jakość badań jest efektem długoletniego doświadczenia i systematycznej praktyki, tak wspomaganego komputerowo analiza danych jakościowych wymaga ciągłego uczenia się, treningu, doskonalenia warsztatu analitycznego i badawczego. Im więcej eksperymentujemy z komputerową analizą danych w badaniach jakościowych, tym bardziej profesjonalny staje się nasz warsztat analityczny. Moje doświadczenia dydaktyczne i doświadczenie w pracy z CAQDAS pokazują, że dużym ułatwieniem w nauce tych programów jest wiedza z zakresu metodologii badań społecznych i analizy danych ilościowych, a także umiejętność pracy z programami do analiz statystycznych³². Badacz, który potrafi sensownie korzystać z tych zasobów, nie tylko uczy się szybciej logiki wspomaganego komputerowo analizy danych jakościowych, ale i potrafi efektywniej korzystać z funkcjonalności programów. Czasami jednak wiedza czy umiejętności tego typu mogą być utrudnieniem w utrzymaniu jakościowego podejścia w analizie danych. W praktyce trudno jest zachować równowagę między tymi podejściami, co skutkuje myśleniem pozytywistycznym o danych jakościowych. Takie myślenie przejawia się przede wszystkim w dążeniu do nadmiernej standaryzacji danych jakościowych – celem wykorzystania technik i procedur znanych z analiz statystycznych.

Doświadczenie w pracy z różnymi programami pozwala także na wskazanie kilku słabości CAQDAS. Jeśli badacz jakościowy pracuje długi czas z jednym wybranym programem CAQDAS, nie zwraca

³² Dla badaczy ilościowych wcześniej pracujących na programach statystycznych specyfika oprogramowania CAQDAS jest łatwiejsza w uczeniu się i rozumieniu logiki wspomaganego komputerowo analizy danych jakościowych.

zwykle szczególnej uwagi na deficyty funkcjonalności. Dopiero możliwość pracy z różnymi programami, porównania ich funkcjonalności uświadamia ich rzeczywistą wartość i użyteczność w badaniach jakościowych. Przykładowo, gdy chcemy analizować duży wolumen danych tekstowych, okazuje się, że programy CAQDAS są często niestabilne i słabo dostosowane do pracy z obszernymi danymi. Stąd w rzeczywistości analiza dużych zbiorów danych wymaga odpowiednich zasobów sprzętowych, najlepiej serwerowych. Słabo są również rozwinięte możliwości przenoszenia i konwersji projektów analitycznych (zakodowanych fragmentów, struktury kodów, map poznawczych itp.) pomiędzy różnymi programami CAQDAS (Corti, Gregory 2011). Brakuje także ujednoliconego formatu przenoszenia projektów pomiędzy programami komercyjnymi. Pomimo że niektórzy dystrybutorzy wprowadzają pewne rozwiązania w tym zakresie (np. zapis projektów analitycznych w standardzie XML), to utrzymanie formatu konwertowanych projektów nadal jest kłopotliwe. Kwestia ta jest szczególnie ważna, gdy chcemy skorzystać w analizie z różnych funkcjonalności występujących w różnych programach. Bardzo często za tymi różnicami, na co zwracałem wcześniej uwagę, stoją metodologie jakościowe, które określają logikę analizy w ramach danego programu. Przykładowo, programom, u podstaw których stoi metodologia teorii ugruntowanej, brakuje dobrej wizualizacji relacji i powiązań zawartych w danych, jakimi cechują się na przykład programy do analizy treści³³. Dla badacza chcącego skorzystać

³³ Doskonałym przykładem jest tu porównanie programów Atlas.ti i QDA Miner. U podstaw rozwoju tego pierwszego leży teoria ugruntowana, zaś drugi oparty jest na technikach analizy treści i przekazów masowych.

z zaawansowanych technik analizy CAQDAS jest to duże utrudnienie. Kwestia ta odnosi się przede wszystkim do programów komercyjnych. Jeszcze innym problemem jest brak możliwości równoczesnej, wieloetapowej i wielopłaszczyznowej pracy grupowej, co uniemożliwia na przykład zespołowe kodowanie w czasie rzeczywistym czy równoczesną analizę danych. Użytkownicy, którzy stykają się z tym problemem, korzystają najczęściej z wersji desktopowych programów CAQDAS. Problem pracy w obrębie zespołu badawczego rozwiązują, wymieniając się pomiędzy sobą plikami z projektami analitycznymi³⁴. Nadal niewielki jest zakres możliwości walidacji danych jakościowych, szczególnie w trakcie procesu kodowania. Nie chodzi tu tylko o analizę rzetelności kodowania, ale o możliwość weryfikacji analizowanych treści. Podobnie z racji tego, że w programach dominuje przede wszystkim metodologia teorii ugruntowanej i analizy treści, trudno jest efektywnie wykorzystać CAQDAS w innych podejściach analitycznych, na przykład w analizie konwersacyjnej czy analizie dyskursu. Brakuje również niektórych technik analizy danych tekstowych, na przykład analiz indukcyjnych i sieciowych.

Użycie CAQDAS nie daje pełnej gwarancji efektywnej analizy danych jakościowych. O tym w dużym stopniu decydują predyspozycje badacza, styl analizy i kodowania danych czy zdolność do myślenia wielowymiarowego. Komputerowa analiza

danych wymaga od badacza cierpliwości i systematyczności, a kodowanie elastyczności, umiejętności wieloaspektowego podejścia do materiału badawczego, szczególnie, gdy kodujemy duży wolumen danych i lista kodowa wymaga ciągłej weryfikacji. W praktyce kodowanie jest tym lepsze, im wyższy jest kapitał lingwistyczny osoby kodującej. Jak już wcześniej wskazywałem, większa wrażliwość i wyższa kompetencja językowa ułatwiają lepsze rozumienie kontekstów znaczeniowych analizowanych danych, a tym samym trafne tworzenie kodów i kategorii. Podobnie umiejętności analityczne czy informatyczne sprzyjają, z jednej strony, lepszej znajomości programów, ich funkcjonalności, jednakże ich przerost – w połączeniu z nadmierną proceduralizacją analizy i rozrostem funkcji analitycznych w samych programach CAQDAS – może prowadzić do zbytnej koncentracji na technice analizy, utrudniając interpretację wyników. Wydaje się, że funkcjonalna elastyczność programów może więc z jednej strony rodzić chaos analityczny lub ograniczać refleksyjność badacza, szczególnie, gdy nie ma on jasno zdefiniowanych celów i zaplanowanych etapów analizy, zaś z drugiej – ułatwiać proces analizy, gdy badacz świadomie korzysta z dostępnych możliwości tkwiących w programach CAQDAS. Jak pokazują ostatnie lata, rozwój CAQDAS to właśnie doskonalenie starych i wprowadzanie nowych funkcjonalności, na przykład automatyczne uczenie wzorców kodowania danych z wykorzystaniem języka NLP (np. program Qualrus) czy zastosowanie *mixed methods* (np. Maxqda czy QDA Miner). Nie jest już także zaskoczeniem korzystanie w analizie danych jakościowych z tabel kontyngencji przy budowaniu typologii czy

klasyfikacji. To implementowanie w programach CAQDAS technik statystycznych jako dodatkowych funkcjonalności może być dla niektórych badaczy zaprzeczeniem istoty badań jakościowych, jednak korzystanie w komputerowej analizie danych jakościowych z technik statystycznych, takich jak analiza korespondencji, skupień, skalowanie wielowymiarowe, regresja czy indukcyjne metody klasyfikacyjne, wzbogaca możliwości eksploracyjne i eksplanacyjne analizy oraz ukazuje nowe pola interpretacji materiału badawczego. Szybki rozwój technologii informacyjnych i do-

stępność dużych wolumenów danych tekstowych i webowych powodują również, że wiele programów CAQDAS podąża obecnie w kierunku analiz online oraz eksploracyjnej metodologii drążenia danych tekstowych – *text mining* (Wiedemann 2013). Niezależnie od tego, jakie funkcjonalności są obecne w programach CAQDAS, badacz jakościowy powinien być na każdym etapie procesu badawczego krytyczny co do efektów stosowania nowych technologii czy funkcjonalności i mieć cały czas na uwadze fakt, że programy CAQDAS są tylko narzędziem w procesie analizy danych.

Bibliografia

Atkinson Paul, Delamont Sara (2009) *Perspektywy analityczne* [w:] Krzysztof Podemski, red., *Metody badań jakościowych. Tom 2*. Warszawa: Wydawnictwo Naukowe PWN, s. 257–284.

Becker Howard S., Gordon Andrew C., LeBailly Robert K. (1984) *Fieldwork with the computer: criteria for assessing systems*. „Qualitative Sociology”, vol. 7, no. 1/2, s.16–33.

Berelson Bernard (1952) *Content analysis in communication research*. Glencoe: Free Press.

Bong Sharon A. (2002) *Debunking Myths in Qualitative Data Analysis*. „Forum Qualitative Sozialforschung”, vol. 3, no. 2 [dostęp 31 marca 2013 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/849>>.

Brent Edward E. (1984) *Qualitative computing: approaches and issues*. „Qualitative Sociology”, vol. 7, no. 1–2, s. 36–60.

Brent Edward E., Anderson Ronald (1990) *Computer Applications in the Social Sciences*. Philadelphia: Temple University Press.

Brosz Maciej (2012) *Komputerowe wspomaganie badań jakościowych. Zastosowanie pakietu NVivo w analizie materiałów nieustrukturyzowanych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 98–125 [dostęp 14 marca 2014 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Brown David (2002) *Going Digital and Staying Qualitative: Some Alternative Strategies for Digitizing the Qualitative Research Process*. „Forum Qualitative Sozialforschung”, vol. 3, no. 2 [dostęp 25 lutego 2013 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/851>>.

Bryda Grzegorz, Anacik Agata, Beata Jaśko (2007) *Ekspertyza efektywności form wsparcia oferowanego długotrwale bezrobotnym w województwie małopolskim – Raport z badań* [dostęp 30 czerwca 2013 r.]. Dostępny w Internecie: <http://www.institut-karpacki.org/index.php?option=com_content&task=view&id=115&Itemid=72>.

Bryda Grzegorz, Anacik Agata, Beata Jaśko (2008) *Badanie potrzeb niepełnosprawnych i oczekiwania pracodawców na rynku pracy – Raport z badań jakościowych* [dostęp 30 kwietnia 2008 r.].

³⁴ Częściowym rozwiązaniem tego problemu jest zakup licencji serwerowej programu CAQDAS. Dystrybutorzy proponują w tym zakresie różne rozwiązania, takie jak m.in. kodowanie wybranych dokumentów, a potem ich łączenie w jeden projekt analityczny czy też przesyłanie jednego pliku projektowego pomiędzy członkami zespołu badawczego.

Dostępny w Internecie: <http://www.institut-karpacki.org/index.php?option=com_content&task=view&id=118&Itemid=72>.

Bryda Grzegorz i in. (2010) *Diagnoza stanu współpracy organizacji pozarządowych w województwie małopolskim – badania jakościowe. Raport z badań* [dostęp 31 października 2010 r.]. Dostępny w Internecie: <<http://www.badaniengo.region-rabka.pl/>>.

Charmaz Kathy (2009) *Teoria ugruntowana. Praktyczny przewodnik po analizie jakościowej*. Przełożyła Barbara Komorowska. Warszawa: Wydawnictwo Naukowe PWN.

Coffey Amanda, Atkinson Paul (1996) *Making sense of qualitative data*. London: Sage.

Conrad Peter, Reinharz Shulamit (1984a) *Caqdas software and qualitative data: editors introductory essay*. „Qualitative Sociology”, vol. 7, no. 1/2, s. 3–15.

Conrad Peter, Reinharz Shulamit (1984b) *Computers and qualitative data*. New York: Human Sciences Press.

Corti Louise, Gregory Arofan (2011) *CAQDAS Comparability. What about CAQDAS Data Exchange?* „Forum Qualitative Sozialforschung”, vol. 12, no. 1 [dostęp 30 maja 2013 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1634>>.

Dey Ian (1993) *Qualitative data analysis: a user-friendly guide for social scientists*. London: New York: Routledge.

Di Gregorio Silvana, Davidson Judith (2008) *Qualitative Research Design for Software Users*. Maidenhead: McGraw Hill.

Drass Kriss A. (1989) *Text analysis and text-analysis software: a comparison of assumptions* [w:] Grant Blank, James L. McCartney, Edward E. Brent, eds., *New in Technology in Sociology: Practical Applications in Research and Work*. New Brunswick: Transaction Publishers, s. s. 155–162..

Fielding Nigel G. (2012) *The Diverse Worlds and Research Practices of Qualitative Software*. „Forum Qualitative Sozialforschung”, vol. 13, no. 2, [dostęp 30 czerwca 2013 r.], Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1845/3369>>.

Fielding Nigel G., Lee Raymond M. (1998) *Computer Analysis and Qualitative Research*. London: Sage Publications.

Fielding Nigel G., Lee Raymond M. (1996) *Qualitative Data Analysis: Representations of a Technology: A Comment on Coffey, Holbrook and Atkinson*. Sociological Research Online, vol. 1, no. 4, [dostęp 21 grudzień 2012], Dostępny w Internecie: <<http://www.socresonline.org.uk/1/4/lf.html>>.

Fielding Nigel G., Lee Raymond M. (1993) *Using Computers in Qualitative Research*. London: Sage Publications.

Fischer Michael D. (1994) *Applications in Computing for Social Anthropologists*. London: Routledge.

Gerson Elihu (1984) *Qualitative research and the computer*. „Qualitative Sociology”, vol. 7, no. 1/2, s. 61–74.

Gibbs Graham (2011) *Analiza danych jakościowych*. Przełożyła Maja Brzozowska-Brywczyńska. Warszawa: Wydawnictwo Naukowe PWN.

Glaser Barney G., Strauss Anselm L. (2009) *Odkrywanie teorii ugruntowanej: strategię badania jakościowego*. Przełożył Marek Gorzko. Kraków: Zakład Wydawniczy Nomos.

Huang Ronggui (2012) *RQDA: R-based Qualitative Data Analysis. R package version 0.2-3*. [dostęp 30 czerwca 2013 r.], Dostępny w Internecie: <<http://rqda.r-forge.r-project.org/>>.

Kelle Udo, ed., (1995) *Computer-aided qualitative data analysis: theory, methods and practice*. London: Sage Publications.

Lewins Ann, Silver Christina (2007) *Using Software in Qualitative Research: A Step-by-Step Guide*. Thousand Oakes: Sage Publications.

Lofland John i in. (2009) *Analiza układów społecznych: przewodnik metodologiczny po badaniach jakościowych*. Przełożyli Sylwia Urbańska, Monika Żychlińska, Anna Kordasiewicz. Warszawa: Wydawnictwo Naukowe Scholar.

Miles Matthew B., Huberman Michael A. (2000) *Analiza danych jakościowych*. Przełożył Stanisław Zabielski. Białystok: Trans Humana.

Niedbalski Jakub (2013) *Odkrywanie Caqdas. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 126–165 [dostęp 14 marca 2014 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Pfaffenberger Bryan (1988) *Microcomputer Applications in Qualitative Research*. Newbury Park: Sage Publications.

Ragin Charles C., Becker Howard S. (1989) *How the microcomputer is changing our analytic habits* [w:] Grant Blank, James L. McCartney, Edward E. Brent Edward, eds., *New Technology in Sociology: Practical Applications in Research and Work*. New Brunswick: Transaction Publishers, s. xx–xx.s. 47–55.

Richards Lyn (2005) *Handling Qualitative Data: A Practical Guide*. London: Sage Publications.

Richards Lyn, Richards Tom (1989) *The impact of computer techniques for qualitative analysis*, Technical Report, no. 6/89. Melbourne: Department of Computer Science, La Trobe University.

Richards Lyn, Richards Tom (1991) *The Transformation of Qualitative Method: Computational Paradigms and Research Processes* [w:] Nigel G. Fielding, Raymond M. Lee, eds., *Using Computers in Qualitative Research*. London: Sage Publications, s. 38–53.

Seale Clive (2001) *Computer-Assisted Analysis of Qualitative Interview Data* [w:] Jaber F. Gubrium, James A. Holstein, eds., *Handbook of Interview Research: Context and Method*. Thousand Oaks: Sage Publications, s. 651–670.

Aneks – oprogramowanie

Ethnograph, <http://www.qualisresearch.com/>

Aquad, <http://www.aquad.de/>

Coding Analysis Software (CAT), cat.ucs.pitt.edu/

HyperRESEARCH, <http://www.researchware.com/products/hyperresearch.html>

Seale Clive (2013) *Wykorzystanie komputera w analizie danych jakościowych* [w:] David Silverman, red., *Prowadzenie badań jakościowych*. Warszawa: Wydawnictwo Naukowe PWN, s. 233–256

Seidel John (1991) *Method and Madness in the Application of Computer Technology to Qualitative Data Analysis* [w:] Nigel G. Fielding, Raymond M. Lee, eds., *Using Computers in Qualitative Research*. London: Sage Publications, s. 107–116.

Silverman David (2007) *Interpretacja danych jakościowych: metody analizy rozmowy, tekstu i interakcji*. Przełożyły Małgorzata Głowacka-Grajper, Joanna Ostrowska. Warszawa: Wydawnictwo Naukowe PWN.

Tesch Renata (1990) *Qualitative Research: Analysis Types and Software Tools*. London, Philadelphia: Falmer Press.

Tesch Renata (1991) *Software for qualitative researchers analysis needs and program capabilities* [w:] Nigel G. Fielding, Raymond M. Lee, eds., *Using Computers in Qualitative Research*. London: Sage Publications, s. 16–37.

The CAQDAS Networking Project – strona projektu [dostęp 30 czerwca 2013 r.]. Dostępny w Internecie: <<http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/>>.

Weitzman Eben A., Miles Mathew B. (1995) *A software source book: Computer Programs for Qualitative Data Analysis*. Thousand Oaks: Sage Publications.

Wiedemann Gregor (2013) *Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences*. „Forum Qualitative Sozialforschung”, vol. 14, no. 2 [dostęp 30 czerwca 2013 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1949>>.

Atlas.ti, <http://www.atlasti.com/product.html>

Maxqda, <http://www.maxqda.com/products>

NVivo, http://www.qsrinternational.com/products_nvivo.aspx

Kwalitan, <http://www.kwalitan.nl/engels/index.html>

Transana, <http://www.transana.org/index.htm>

Qualrus, <http://www.ideaworks.com/qualrus/>

TAMS Analyzer, <http://tamsys.sourceforge.net>

QDA Miner, <http://provalisresearch.com/products/>

Xsight (QSR), http://www.qsrinternational.com/products_xsight.aspx

EdEt (Edytor etnograficzny), <http://www.etnologia.uw.edu.pl/etno/dlaStudentow/edet>

Weft QDA, <http://www.pressure.to/qda/>

RQDA (R software), <http://rqda.r-forge.r-project.org>

Saturate, <http://www.saturateapp.com>

CATMA, <http://www.catma.de/versions>

Compendium, <http://compendium.open.ac.uk/institute/download/download.htm>

Qiqqa, <http://www.qiqqa.com>

Digital Replay Systems (DRS), <http://thedrs.sourceforge.net>

Dedoose, <http://www.dedoose.com>

LibreQDA, <http://www.libreqda.edu.uy/en>

HyperQual Lite (HyperQual), <http://hyperqual.on-rev.com/hyperqual/>

Cytowanie

Bryda Grzegorz (2014) *CAQDAS a badania jakościowe w praktyce*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 12–38 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

CAQDAS and Qualitative Research Practice

Abstract: The purpose of this article is methodological reflection on dialectical relationship between the qualitative research process and the process of computer-assisted qualitative data analysis. Basing on many years of experience in the work with various CAQDAS tools, the author tries to show the impact they have on shifting the way of thinking on methodology, the process of data analysis and conducting qualitative research. The specificity of CAQDAS usage in research practice requires methodological rigor in the process of collection and archiving of data, as well as the accuracy and precision in the process of coding, analyzing, and visualizing data. The use of computer-aided analysis of qualitative data in research practice not only shapes a framework for the sociological interpretation but also changes the way of perceiving research problems. The essence of this process is a specific interaction between new technologies and traditional methodology, data analysis, and qualitative research. In this sense, the use of CAQDAS in qualitative research practice is shaping the personality and identity of qualitative researcher, his/her style of work, data analysis, and conducting the fieldwork; it develops his or her new analytical and computer skills without which it is difficult to imagine a modern qualitative sociological research.

Keywords: CAQDAS, information technology, data analysis, qualitative research, style of data analysis, mixed methods

Maciej Brosz

Uniwersytet Gdański

DOI: <https://doi.org/10.18778/1733-8069.10.2.03>

Badanie jakościowe w warunkach niedoszacowania czasu – czyli o konieczności „podążania na skrót”. Organizacja procesu badawczego pod kątem zastosowania komputerowego wsparcia analizy danych jakościowych

Abstrakt Artykuł poświęcony jest organizowaniu procesu badawczego pod kątem zastosowania oprogramowania NVivo. Prezentowane badania zostały przeprowadzone w warunkach niedoszacowania czasu. Zastosowanie CAQDA oraz odpowiednia organizacja badania pozwoliły zmieścić się w harmonogramie narzuconym przez instytucję zlecającą. Organizacja i podział pracy, przygotowanie odpowiednio sformatowanych dokumentów oraz uzupełnienie jakościowej bazy danych o zbiór atrybutów bazujących na danych standaryzowanych znacząco usprawniły opracowywanie wyników.

Słowa kluczowe CAQDAS, NVivo, autokodowanie, segmentacja danych, importowanie ilościowych baz danych, kwerenda macierzowa, macierz ramowa

Maciej Brosz, asystent w Zakładzie Socjologii Stosowanej Instytutu Filozofii, Socjologii i Dziennikarstwa Uniwersytetu Gdańskiego oraz analityk w Pracowni Realizacji Badań Socjologicznych UG. Jego główne zainteresowania to socjologia mieszkalnictwa, rewitalizacja zdegradowanych obszarów miejskich, problematyka bezrobocia oraz metodologia nauk społecznych.

Adres kontaktowy:

Instytut Filozofii
Socjologii i Dziennikarstwa
Uniwersytet Gdański
ul. Bażyńskiego 4, 80-952 Gdańsk
e-mail: maciej.brosz@ug.edu.pl

Niniejszy artykuł poświęcony jest, po pierwsze, prezentacji zastosowania programu NVivo jako kluczowego elementu warsztatu metodologicznego w badaniach jakościowych prowadzonych w warunkach wąskich ram czasowych, po drugie – podejmuje kwestię projektowania procesu badawczego pod kątem zastosowania komputerowego wsparcia analizy danych jakościowych (CAQDA), po trzecie, wskazuje na istotny problem wynikający z rosnących mocy obliczeniowych komputerów osobistych i tym samym możliwości oprogramowania CAQDA.

Autor artykułu jest świadomy faktu, że niektóre z omawianych tu procedur powinny zostać opisa-

ne z pominięciem skrótów myślowych i z pełnym uzasadnieniem poszczególnych stwierdzeń. W ten sposób tekst stałby się bardziej użyteczny dla badaczy, którzy rozpoczynają swoje poszukiwania w obszarze metod komputerowego wspomaganie badań jakościowych. Stosowanie takiego podejścia nie jest jednak konieczne, ponieważ w ostatnich latach opublikowano szereg tekstów, które przybliżają podstawowe zagadnienia związane z CAQDA (zob. np. Bieliński, Iwińska, Rosińska-Kordasiewicz 2007; Brosz 2012; Niedbalski, Ślęzak 2012; Niedbalski 2013). Zasadne wydaje się zatem pominięcie ich w niniejszym opracowaniu przy jednoczesnym zogniskowaniu uwagi na kwestiach bardziej złożonych.

Logika funkcjonowania rynku badawczego oraz projektowe realia prowadzenia działalności badawczej wyznaczają określony reżim czasowy. Na ogół oznacza to konieczność przeprowadzania „szybkich” badań – najlepiej „na wczoraj” lub też z nieodpowiedzialnie zaplanowanym kalendarzem działań. Nowoczesne technologie znajdujące wyraz w najnowszych wydaniach specjalistycznego oprogramowania dostarczają funkcjonalności – w połączeniu z mocami obliczeniowymi komputerów – umożliwiające stawianie czoła wymogom współczesności.

Prowadzenie działalności badawczej w ramach przedsięwzięć i projektów finansowanych ze środków unijnych związane jest nierzadko z doświadczaniem przez badaczy konsekwencji wdrażania założeń przyjmowanych przez podejmujących działania w ramach projektów. Założenia te, przyjmowane często bez konsultacji ze specja-

listami zajmującymi się działalnością badawczą, dotyczą bardzo istotnych zagadnień. Przykładem mogą tu być: harmonogram realizacji badań, metodologia badań oraz dobór technik gromadzenia informacji, wielkości próby i inne. Nie jest celem niniejszego artykułu przedstawienie owych kłopotliwych konsekwencji, ale raczej strategii radzenia sobie z niektórymi z nich. W tym konkretnym przypadku chodzi o konsekwencje przewidzianych w projekcie – i tym samym wpisanych do wniosku projektowego – ram czasowych oraz liczebności wywiadów. Ostatnim ogniwem w łańcuchu decyzyjnym okazał się wybrany podmiot przeprowadzający badanie, co oznaczało konieczność postępowania według podjętych wcześniej założeń. Wyzwanie okazało się niebagatelne: osiemdziesiąt indywidualnych wywiadów pogłębianych, z czego trzydzieści w trudnodostępnej zbiorowości, a do tego – jeden miesiąc na realizację fazy terenowej i opracowanie raportu. Należy w tym miejscu zaznaczyć, że – choć z dużym wysiłkiem – badanie zostało zrealizowane. Zgromadzone przy tej okazji doświadczenia stanowią podstawę rozważań o możliwościach współczesnych programów komputerowych oraz o potencjalnych skutkach podążania „na skrót” w badaniach jakościowych.

Na przestrzeni ostatniej dekady moc obliczeniowa standardowego komputera osobistego wzrosła kilkakrotnie, potencjał ten został efektywnie wykorzystany przez twórców oprogramowania CAQDA. Przykładowo – w 2004 roku, chcąc przeprowadzić badanie jakościowe ze wspomaganie komputerowym obejmującym tekstualizację danych, kodowanie, wizualizację powiązań pomiędzy

kategoriami oraz analizę materiałów tekstowych pod względem ilościowym, badacz musiałby zaopatrzyć się w kilka programów. Obecnie do tego celu wystarczy jeden. Nie wyklucza to oczywiście sytuacji, w której jedne narzędzia są przeznaczone do bardziej specyficznych zastosowań niż inne. Niemniej jednak wszystkie charakteryzują się znaczną wszechstronnością. Program z rodziny CAQDA oferujący różnorodne możliwości to złożone narzędzie, do którego użycia niezbędne jest zapoznanie się z jego możliwościami, a następnie wdrożenie ich do prowadzonej działalności. Paradoksalnie, studia poświęcone posługiwaniu się programem oraz zastosowaniu jego możliwości w pracy badawczej mogą zająć tyle czasu, ile mogłyby zająć studiowanie złożonego zjawiska społecznego.

Nowoczesne narzędzia komputerowego wsparcia analizy jakościowej są tworzone zgodnie z zasadą przyjazności dla użytkownika (*user friendly*). Oznacza to, że przeprowadzenie obróbki i eksploracji danych – automatyczne kodowanie bazujące na słowach kluczowych, wprowadzenie tak zwanej *stoplisty*, polecenie policzenia ilościowych wskaźników ilustrujących zależności w obrębie materiałów tekstowych, a następnie ich wizualizacja w postaci klastrów, dendrogramów, map 3D – sprowadza się do kilku kliknięć przyciskami myszki. Złożona aparatura wspomagająca proces analizy może w takiej sytuacji stać się rodzajem „czarnej skrzynki”, która *coś* robi z danymi, ale nie do końca wiadomo co. Z jednej strony potencjał narzędzi CAQDA pozwala efektywniej gospodarować czasem, z drugiej zaś – nakłada na badacza obowiązek kontrolowania procesu przetwarzania

danych. Kompetencje niezbędne do sprawowania takiej kontroli wymagają konkretnych umiejętności charakterystycznych dla programistów i informatyków, jak na przykład pisanie skryptów w określonym języku programowania. Nie są to umiejętności powszechnie występujące w środowisku socjologów, antropologów czy szerszej – humanistów. Umieszczone w tytule artykułu wyrażenie „podążanie na skrót” ma za zadanie poznać zaniepokoić zarówno zwolenników, neutralnych użytkowników, jak i krytyków stosowania w badaniach jakościowych wspomaganie komputerowego. Niepokój ten powinien uwrażliwić na towarzyszącą rozwojowi oprogramowania i jego wszechstronności potrzebę selektywnego – a tym samym przemyślanego i kontrolowanego – wykorzystywania poszczególnych funkcjonalności. To, że narzędzie jest wszechstronne, nie stanowi przesłanki do użycia wszystkich jego możliwości do analizy każdego materiału. Podobnie jak użytkownik programu SPSS, który ma do wyboru kilkadziesiąt testów statystycznych oraz narzędzi przetwarzania bazy danych, tak i operator CAQDA powinien uruchamiać poszczególne procesy w gruntownie przemyślanym i zaplanowanym porządku metodologicznym.

Prezentowane w niniejszym artykule rozwiązania mają za podstawę badania jakościowe o charakterze diagnostycznym, przeprowadzone w lipcu i sierpniu 2013 roku przez Pracownię Realizacji Badań Socjologicznych Uniwersytetu Gdańskiego we współpracy z Regionalnym Centrum Wolontariatu w Gdańsku. Przedsięwzięcie badawcze zostało przeprowadzone w ramach projektu „Gdański model wolontariatu seniorów – krok 1”, finansowa-

nego z Rządowego Programu na rzecz Aktywności Społecznej Osób Starszych na lata 2012–2013 (Kiwi 2013). Głównym celem przeprowadzonego badania jakościowego było rozpoznanie opinii dwóch zróżnicowanych grup respondenckich: przedstawicieli organizacji pozarządowych i instytucji publicznych oraz liderów oddolnych inicjatyw, czyli aktywnych seniorów działających jako wolontariusze na terenie Gdańska. W badaniu poruszono kwestie szans, barier, metod i obszarów animowania współpracy obu środowisk. Praktycznym celem podjętych badań było wypracowanie rekomendacji ułatwiających projektowanie działań wspierających oddolne inicjatywy wolontariackie wśród seniorów (osób powyżej 65. roku życia).

W ramach projektu założono przeprowadzenie łącznie osiemdziesięciu indywidualnych wywiadów pogłębionych, w tym pięćdziesięciu z przedstawicielami organizacji i instytucji oraz trzydziestu wśród liderów, seniorów – wolontariuszy. Dobór respondentów do próby badawczej miał charakter celowy i został oparty na operatach sporządzonych przez Regionalne Centrum Wolontariatu w Gdańsku na podstawie: po pierwsze, bazy organizacji samorządowych i instytucji publicznych, prowadzonej przez Wydział Polityki Społecznej Miasta Gdańsk oraz, po drugie, bazy inicjatyw senioralnych, klubów seniora i uniwersytetów trzeciego wieku, prowadzonej przez Pełnomocnika Prezydenta Miasta Gdańska ds. Seniorów. W toku badań terenowych, ze względu na trudności z uzyskaniem dostępu (okres urlopowy), udało się zrealizować czterdzieści dziewięć wywiadów z przedstawicielami oraz dwadzieścia jeden z liderami.

Do głównych obszarów tematycznych przeprowadzonych wywiadów należały:

- prowadzona działalność na rzecz seniorów lub we współpracy z wolontariuszami – seniorami,
- uwarunkowania upowszechnienia się wolontariatu osób starszych,
- sposoby i skuteczność docierania do seniorów z informacjami o możliwościach podejmowania działań,
- problemy i potrzeby osób starszych podejmujących się działań oddolnych i angażujących się w działalność wolontaryjną,
- potrzeby edukacyjne osób starszych oraz współpracujących z osobami starszymi w zakresie posiadanych wiadomości, umiejętności i kwalifikacji.

Na potrzeby projektu przygotowano scenariusze wywiadów zróżnicowane formalnie pod względem specyfiki dwóch grup respondenckich objętych badaniem. Celem tego zabiegu było uzyskanie oglądu zjawisk z dwóch różnych perspektyw: osób reprezentujących organizacje pozarządowe lub instytucje publiczne oraz liderów. Projektując procedurę badawczą, uwzględniono specyfikę prowadzenia wywiadu z osobą starszą, czyli stan fizyczny, możliwości percepcyjne oraz konieczność dostosowania tempa rozmowy i używanego słownictwa do konkretnego rozmówcy (zob. Błędowski, Mossakowska, Więcek 2012; Czekanowski 2012; Czekanowski, Załęcki, Brosz 2013).

Jak wyżej wspomniano, badania terenowe były przeprowadzane w krótkim przedziale czasu – niedoszacowanym przez wnioskodawców. Kluczowymi elementami strategii badawczej były: po pierwsze, zorganizowanie prac terenowych, minimalizujące liczbę i zakres czynności związanych z obróbką danych, przy maksymalnym wykorzystaniu możliwości programu NVivo z zakresu automatyzacji procesu porządkowania danych (prezentowane tu rozwiązania odnoszą się do wersji programu 9.2 i nowszych), po drugie – efektywne wykorzystanie czasu poprzez przesunięcie standaryzowanych operacji klasyfikowania materiału na osoby prowadzące wywiady, przy wykorzystaniu funkcji formatowania tekstu i atrybutyzacji poszczególnych przypadków (program NVivo udostępnia funkcje, które umożliwiają automatyczne kodowanie tekstów o określonym formacie oraz przypisywanie przypadkom atrybutów na podstawie macierzowych baz danych). Zastosowanie oprogramowania CAQDA miało spowodować redukcję czasu pracy analityka, polegającą na poddaniu analizie jakościowej (w uproszczeniu: lektura, kodowanie, typologizowanie, porównywanie, formułowanie wniosków) wyłącznie tych wypowiedzi, które zostały wyartykułowane przez informatorów posiadających faktyczne doświadczenia związane z danym zagadnieniem. Zabieg, który zastosowano, można określić mianem segmentacji mającej na celu zwiększenie prawdopodobieństwa uzyskania dostępu do informacji wysycających problematykę badań. Powodzenie badań opartych na takich rozwiązaniach nie byłoby jednak możliwe, gdyby nie staranne zaprojektowanie całego przedsięwzięcia od początku do końca. W dalszej części tekstu zostaną zaprezentowane kluczowe

elementy procesu badawczego z użyciem oprogramowania CAQDA.

Przegląd poszczególnych działań zaczniemy od tego, co najważniejsze i stanowiące „elementarz” badań socjologicznych *sensu largo*: podczas spotkania z klientem ustalono cele i założenia projektu pozwalające sformułować problematykę, zakres i formę przetwarzanych w badaniu i prezentowanych w raporcie informacji. Ze względu na ograniczone ramy czasowe, spotkania te przybrały formę negocjacji, których celem było wypracowanie kompromisu godzącego oczekiwania uczestników projektu z możliwościami podmiotu realizującego badanie. Uzgodnienia poczynione w tym kluczowym stadium stanowiły podstawę realizowanych w kolejnych etapach projektu działań i w zasadzie stanowiły warunek *sine qua non* podjęcia wyzwania. Rezultatem owych ustaleń był harmonogram prac zespołu przeprowadzającego badanie.

Kolejnym etapem pracy było przygotowanie narzędzia badawczego. Ze względu na eksploracyjny charakter badań zaplanowano wykorzystanie techniki indywidualnego wywiadu pogłębionego. Przygotowania w tym zakresie nie ograniczyły się tylko i wyłącznie do opracowania dyspozycji odpowiadających problematyce badania. Strategiczny zamiar zredukowania czasochłonnych elementów procesu wymagał – już na etapie budowy narzędzia – uwzględnienia specyfiki programu komputerowego jako narzędzia opracowywania materiału. Tak jak w arkuszach kalkulacyjnych (m.in. SPSS, Statistica czy Excel) informacje są utrwalane w postaci dwuwymiarowych macierzy zawierających w wierszach rekordy, a w kolumnach zmienne, tak i każdy

program CAQDA charakteryzuje się określonym sposobem przechowywania danych. Innymi słowy, program CAQDA to szczególnego rodzaju nośnik, którego kształt i specyfika kształtują rejestrowaną informację. Wyrażenie informacji w określonej formie odsyła do rozważań poświęconych zagadnieniu pomiaru w naukach społecznych. Znane socjologom kategorie poziomów pomiaru: nominalny, porządkowy, interwałowy i ilorazowy w istocie oznaczają to, jak rejestrowana jest określona informacja (Nowak 2007). Spoczywająca na badaczu konieczność uwzględnienia specyfiki nośnika, jakim jest stosowany program CAQDA (np. NVivo, Atlas.ti, QDA Miner, OpenCode, RQDA + środowisko R i inne), nadaje problematyce pomiaru praktyczny wymiar. Użytkownicy planujący zastosowanie CAQDA w badaniach powinni w pierwszej kolejności zapoznać się ze sposobem „czytania” danych przez narzędzie. Nie chodzi tu wyłącznie o format pliku komputerowego, ale znaczenie – na gruncie danego programu – takich terminów, jak źródło, kod, kategoria, nota, kwerenda, atrybut oraz zależności między nimi. W każdym z programów pojęcia te mają odmienne desygnaty, co może utrudniać korzystanie z produktów różnych firm, jak i komplikować komunikację między użytkownikami różnych narzędzi (Saillard 2011), nie wspominając o krzyżowej konwersji danych pomiędzy różnymi programami.

Program NVivo pozwala na import plików o różnych formatach, kategoryzowanie treści na wiele sposobów oraz przypisywanie poszczególnym źródłom, przypadkom, kategoriom kodowym atrybutów (Brosz 2012; Niedbalski, Ślęzak 2012). Prezentacja tych operacji nie jest przedmiotem niniejszego artykułu. Algorytm budowy bazy danych obej-

mował następujące działania: po pierwsze, import przepisanych wywiadów do programu i nadanie im statusu przypadków, po drugie – przypisanie poszczególnym przypadkom atrybutów – zmiennych zawierających informacje standaryzowane – umożliwiającymi segmentowanie/agregowanie zbioru, po trzecie, czyszczenie bazy danych.

Uwzględnienie w projekcie narzędzia zróżnicowania osób biorących udział w badaniu na dwie zbiorowości – przedstawicieli podmiotów instytucjonalnych oraz liderów inicjatyw lokalnych – wymagało opracowania pasujących do siebie zestawów dyspozycji odpowiadających typowi rozmówcy. W praktyce te same zagadnienia musiały zostać wyrażone w sposób dopasowany do wywiadu z liderem, osobą starszą (rys. 1) i przedstawicielem instytucji publicznej (rys. 2).

Jak widać na załączonych ilustracjach (rys. 1 i 2), zagadnienia poruszane na początku wywiadu różniły się w zależności od typu rozmówcy. O ile wywiad z przedstawicielami instytucji publicznych można było rozpocząć od względnie konkretnego pytania o specyfikę działalności, to rozmowa z osobą starszą wymagała odpowiedniego przygotowania. Rozdzielenie pytania 1 na część 1A i 1B związana była z potrzebą wprowadzenia rozmówcy w obszar tematyczny badania. Pojęcia takie jak „wolontariat”, „działanie na rzecz innych”, „działalność społeczna” mogły nie być przez badanych kojarzone z prowadzonymi na co dzień działaniami. Obok zagadnień uszczegóławiających, pozwalających elastycznie prowadzić rozmowę, poszczególne dyspozycje (w każdym zestawie było ich po osiem) zawierały też pomocniczą listę informacji do ustalenia.

Rys. 1. Fragment dyspozycji do wywiadu z liderami

Na początku wywiadu nie używamy określeń: „wolontariat” + pochodnych, „niesienie pomocy innym”.

1. Proszę opowiedzieć mi o swojej działalności... na rzecz innych/w dzielnicy/w ramach...

1A. Jak to się zaczęło? Ktoś Pana(nią) zainspirował, może coś zainspirowało? W jaki sposób rozpoczął(ęła) Pana(i) działać na rzecz/w sprawie?

Ta część rozmowy pełni funkcje wprowadzające, chodzi o uchwycenie tła biograficznego, rodzinnego, genealogicznego – respondent ma się odnaleźć i zdefiniować siebie w temacie. Zadaniem osoby prowadzącej wywiad jest spowodowanie urefleksyjnienia działalność badanego(ej), tak aby mógł/mogła o niej opowiadać. Chodzi o to, aby badany sam zdefiniował to co robi.

Co należy ustalić:

- czy to są osoby aktywne w swoim życiu od zawsze, czy też takie, które zaktywizowały się po zakończeniu aktywności zawodowej, kiedy np. pojawił się czas do zagospodarowania
- jakie były drogi dojścia do działalności, aktywności (wolontaryjnej)

1B. Na czym polega Pana(i) działalność? Co Pan(i) robi, jaką rolę wypełnia/odgrywa/realizuje...

Chodzi zarówno o przedsięwzięcia, w które włączał(a) się Pan(i) jako uczestnik(-czka), jak i jako inicjator(ka), organizator(ka) lub w zbliżonej roli.

- w roli odbiorcy oferty – uczestnika
- w roli inicjatora, organizatora – lidera, kreatora,
- w roli osoby wspomagającej przebieg przedsięwzięcia – np. wolontariusz, wolontariusz zadaniowy, pomocnik
- a może jest to jakaś zupełnie inna rola, jaka...**

Czy działania, które Pan(i) opisał(a) miały charakter jednorazowy czy też cykliczny, powtarzalny (jednorazowe czy też z jakimś ciągiem dalszym)?

Jeśli charakteryzowały się jakąś regularnością, to od czego ona zależała? (sezonowość/pora roku, umiejętności, majątność, samopoczucie... czy coś innego)

Jakie są Pana(i) doświadczenia, jako osoby uczestniczącej w programach, przedsięwzięciach, w których brały osoby starsze? Jaka jest Pana(i) opinia o wolontariacie osób starszych?

Co należy ustalić:

- czy badany(a) faktycznie był(a)/jest wolontariuszem
- czy rozmówca brał udział w przedsięwzięciach adresowanych do osób starszych
 - co osoba robiła jako wolontariusz
 - jeśli osoba działała w ramach organizacji to czy dostrzega jakieś niedostatki jej funkcjonowania – czego należałoby uczyć tych, którzy pracują w takich organizacjach
 - czy korzystała z jakichś form wsparcia adresowanych dla inicjatorów działań
 - czy i co osoba robiła jako wolontariusz na rzecz seniorów
- jak oceniana jest działalność seniorów w ramach wolontariatu

Źródło: opracowanie własne.

Rys. 2. Fragment dyspozycji do wywiadu z przedstawicielami instytucji publicznych

1. Proszę opowiedzieć mi o realizowanych przez Państwa przedsięwzięciach, na rzecz..., przy udziale... osób starszych, czyli uwzględniając osoby starsze jako odbiorców działań lub ich współorganizatorów czy wolontariuszy.

W pierwszym rzędzie proszę odnieść się do tych działań, w które są lub mogliby być zaangażowani wolontariusze. Chodzi zarówno o obecność seniorów w roli odbiorców, adresatów oferty, jak i osób działających w roli wolontariuszy (lub zbliżonej roli).

- w roli odbiorców oferty - ci, do których kierowana jest oferta
- w roli osób działających na rzecz - "świadczący usługi", działający na rzecz...

Jakie są Państwa doświadczenia we współpracy z osobami starszymi jako wolontariuszami? Jaka jest Pana(i) opinia o wolontariacie osób starszych?

Co należy ustalić:

- w kwestii wolontariatu:
 - co jest – co mogłoby być
 - jeśli nie ma – dla czego nie ma
- czy organizacja/instytucja korzysta z wolontariatu w ogóle i z wolontariatu osób starszych
 - co musiałyby się stać, aby w realizowanych działaniach ująć wolontariuszy z grupy 60+
- czy w tej organizacji/instytucji istnieje oferta dla seniorów
 - czy i co organizacja/instytucja robi przy udziale seniorów (jako wolontariuszy)
 - jakie formy wsparcia są stosowane
 - czy i co organizacja/instytucja robi na rzecz seniorów
- jak oceniana jest działalność seniorów w ramach wolontariatu

Źródło: opracowanie własne.

Indywidualne wywiady pogłębione, prowadzone w oparciu o przygotowane dyspozycje, były nagrywane. Do zadań osób prowadzących wywiady należało także sporządzenie transkrypcji. W tym celu przygotowano specjalne arkusze transkrypcyjne, w których zastosowano wzór formatowania tekstu w zakresie przypisania określonego stylu (np. „Nagłówek 1”, „Nagłówek 2” lub „Normalny”). Rozwią-

zanie to umożliwiło wykorzystanie narzędzia autokodowania w programie NVivo. Zamieszczona poniżej ilustracja (rys. 3) przedstawia fragment arkusza transkrypcji wywiadu z liderami (arkusze i inne narzędzia miały swoje odpowiedniki w odniesieniu do przedstawicieli i liderów – ich graficzna forma nie różniła się, dlatego dalsza prezentacja opiera się na materiałach dotyczących liderów).

Rys. 3. Fragment arkusza transkrypcji wywiadu z liderem

transkrypcja wywiadu nr []

Q1. Proszę opowiedzieć mi o swojej działalności... ← nagłówek 1

Q1A. Jak to się zaczęło? ← nagłówek 2

piszemy wyłącznie w obrębie szarych pól ← tekst podstawowy

Q1B. Na czym polega Pana(i) działalność? Co Pan(i) robi, jaką rolę wypełnia/odgrywa/realizuje...

[]

o przyjmowanej roli

[]

powtarzalność, cykliczność, regularność

[]

doświadczenia udziału w przedsięwzięciach, w których brały osoby starsze

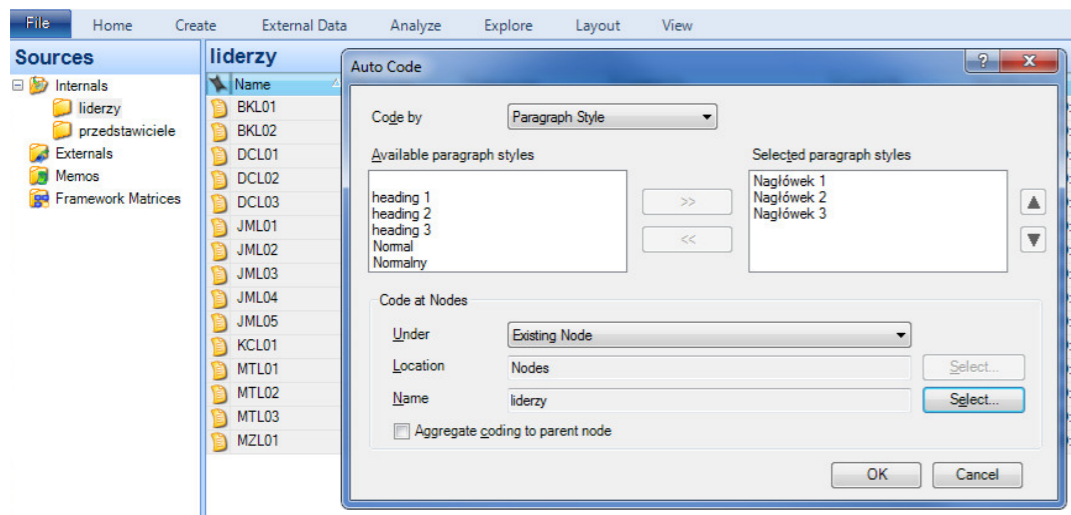
[]

Źródło: opracowanie własne.

Jak łatwo się domyślić, oznaczenia: Q1, Q1A i Q1B odnoszą się do poszczególnych dyspozycji. Dokument został hierarchicznie sformatowany w edytorze tekstu (dowolny: MS Word, OpenOffice.org – Writer, LibreOffice – Writer lub inny z funkcją formatowania) według zasady: główne zagadnienie jako nagłówek 1, zagadnienie podrzędne jako nagłówek 2 i tak

dalej. Pole o formacie tekstu podstawowego zostało zaznaczone szarym kolorem i zdefiniowane jako obszar, w którym osoba przepisująca miała wpisać fragmenty odpowiadające poszczególnym tematom. Ustrukturyzowany w powyższy sposób wywiad można zaimportować do programu NVivo, a następnie poddać procesowi autokodowania.

Rys. 4. Autkodowanie dokumentów zaimportowanych do NVivo



Źródło: opracowanie własne.

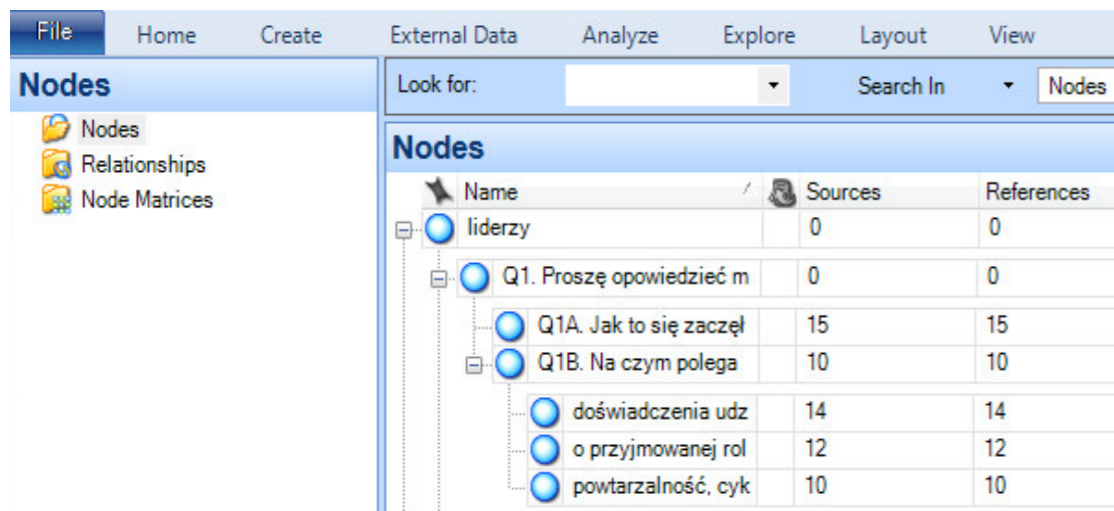
Zabieg ten spowodował powstanie struktury kategorii kodowych odpowiadającej hierarchii zastosowanych formatów (patrz rys. 5).

selektywnego oglądu danych i podjąć prace analityczne. Możliwości programu są jednak większe. Warunkiem jest skorzystanie z funkcji atrybutyzacji.

W taki oto sposób badacz dysponuje materiałem uporządkowanym według założonych kategorii klasyfikacyjnych. Już na tym etapie można dokonywać

Technika indywidualnego wywiadu pogłębionego umożliwia rejestrowanie informacji niestandardowych, wymykających się klasyfikacjom, typolo-

Rys. 5. Struktura kategorii kodowych po procesie autkodowania



Źródło: opracowanie własne.

giom, kafeteriom budowanym przez badacza. Posługiwanie się informacjami o typowym dla wywiadu jakościowego charakterze (narracyjność, opisowość, rozbudowanie itp.) w fazie porządkowania materiału byłoby niezwykle czasochłonne. Każdy z osiemdziesięciu przypadków musiałby bowiem zostać zakodowany odpowiednimi znacznikami. Stosując narzędzia z rodziny CAQDA, uzyskanie dostępu do informacji pozwalających grupować, agregować i sortować poszczególne przypadki można znacząco usprawnić. Należy w tym miejscu zaznaczyć, że zabieg ten nie ingeruje w jakościowy charakter techniki wywiadu pogłębionego. Zastosowane rozwiązanie polegało na przygotowaniu zestawu standaryzowanych zmiennych dychotomicznych („tak”/„nie” lub „posiada cechę”/„nie posiada cechy”), które mogłyby być z powodzeniem rejestrowane przez prowadzącego wywiad w czasie jego trwania (zaraz

po lub podczas dokonywania transkrypcji) i nanoszone do przygotowanej bazy danych. Zastosowanie zmiennych na innych poziomach pomiaru niż nominalny również byłoby możliwe, ale mogłoby niepotrzebnie skomplikować pracę podczas wywiadu. Wielokrotnie wspomniane w tym artykule ograniczenia czasowe skłoniły autora badania do zastosowania procedury „opisania” poszczególnych wywiadów atrybutami zawierającymi standaryzowane informacje. Oprócz scenariuszy wywiadów zawierających dyspozycje, prowadzący zostali wyposażeni w specjalne arkusze. Można tu użyć określenia kwestionariusze, bowiem obok stwierdzeń zawierały one pytania do rozstrzygnięcia oraz dane metryczkowe (rys. 6). Forma pytań, stwierdzeń oraz zastosowane warianty odpowiedzi miały z założenia upraszczać i sprowadzać dane do funkcji atrybutów, które można przypisać przypadkom.

Rys. 6. Fragment arkusza informacji standaryzowanych do wywiadu z liderami

Pytania dodatkowe do wywiadu liderem, wolontariuszem (kwestionariusz wypełnia prowadzący wywiad)	oznaczenie wywiadu		
	tak	nie	nie dotyczy/brak informacji
1. Inspiracją do podjęcia były doświadczenia rodzinne (ktoś w rodzinie działał...)	1	2	3
2. Inspiracją do aktualnej działalności było „bycie aktywnym od zawsze”	1	2	3
1 3. Respondent jest „aktywny od zawsze”	1	2	3
A 4. Respondent jest aktywny odkąd jest na emeryturze i ma „więcej czasu” UWAGA: należy pamiętać, że osoby aktywne i na emeryturze, często mówią: „nie mam na nic czasu”	1	2	3
1. czy badany(a) faktycznie był(a)/jest wolontariuszem	1	2	3
2. czy rozmówca brał udział w przedsięwzięciach adresowanych do osób starszych	1	2	3
3. czy korzystała z jakichś form wsparcia adresowanych dla inicjatorów działań	1	2	3
4. czy osoba robiła coś jako wolontariusz na rzecz seniorów	1	2	3
1 5. czy zamierza w przyszłości udzielać się jako wolontariusz	1	2	3
B 6. czy brał udział w przedsięwzięciach adresowanych dla osób starszych	1	2	3
1. jako uczestnik (korzystający, konsument)	1	2	3
2. jako organizator	1	2	3
3. jako wspomagający, wolontariusz, pomocnik	1	2	3
4. inna rola	1	2	3
7. czy osoba przynależy do jakiejś organizacji lub stowarzyszenia	1	2	3

Źródło: opracowanie własne.

Opracowanie wywiadu po jego przeprowadzeniu polegało na sporządzeniu transkrypcji oraz wpisaniu informacji standaryzowanych do przygotowanej bazy danych (rys. 7). Baza ta nie różniła się niczym od typowej, stosowanej w programach

komputerowych wykorzystywanych w badaniach ilościowych. Numeracja pytań miała swoje odpowiedniki w nazwach zmiennych – zadanie osoby kodującej polegało na wpisaniu wartości od 1 do 3 do odpowiedniej komórki arkusza kalkulacyjnego.

Rys. 7. Fragment bazy danych przygotowanej do wpisywania informacji

oznaczenie wywiadu	Q1A.1	Q1A.2	Q1A.3	Q1A.4	Q1B.1	Q1B.2	Q1B.3

Źródło: opracowanie własne.

Po wykonaniu swojego zadania osoby prowadzące wywiady i opracowujące je według nakreślonych powyżej zasad przekazywały pliki komputerowe zawierające transkrypcje poszczególnych wywiadów, bazę danych oraz pliki audio zawierające nagrania. Transkrypcje importowano do NVivo, zaś bazy danych zostały scalone i poddane drobnym obróbkom, aby zawarte w nich klasyfikacje były czytelne w programie. Wpisując informacje do bazy,

„koderzy” używali wartości liczbowych (od 1 do 3), co znacząco ułatwiło i przyspieszyło pracę (łatwiej wpisywać pojedyncze cyfry niż całe słowa – „tak”, „nie”). Podczas pracy w NVivo taka postać danych byłaby utrudnieniem. Przetworzenie całej bazy polegało na użyciu funkcji „znajdź i zamień, zamień wszystkie” w arkuszu kalkulacyjnym. I tak, „1” zamieniono na „tak”, „2” na „nie”, a „3” na akronim „ND/BI” (nie dotyczy / brak informacji).

Rys. 8. Fragment bazy danych – gotowy do importu do programu NVivo

atr.liderzy	Q1A.1	Q1A.2	Q1A.3	Q1A.4	Q1B.1	Q1B.2	Q1B.3
BKL01	nie	tak	tak	nie	tak	tak	tak
BKL02	nie	nie	tak	tak	nie	tak	tak
DCL01	tak	nie	tak	tak	tak	tak	tak
DCL02	tak	tak	tak	tak	tak	tak	tak
DCL03	nie	nie	nie	nie	tak	tak	tak
DCL04	tak	tak	nie	nie	tak	tak	ND/BI
JML01	nie	tak	nie	tak	tak	tak	tak
JML02	nie	tak	tak	nie	tak	tak	tak
JML03	tak	tak	tak	nie	tak	ND/BI	tak
JML04	nie	nie	nie	tak	tak	tak	tak
JML05	nie	nie	nie	tak	tak	tak	nie
KCL01	nie	tak	tak	tak	tak	tak	nie

Źródło: opracowanie własne.

Aby połączyć macierz danych z przypadkami w NVivo, należy spełnić kilka warunków. Konieczne jest tu zwrócenie uwagi na specyficzne nazewnictwo. Przypadkiem (*Case*) nazwiemy wywiad lub serię wywiadów z określonym badaniem, na przykład BKL01, DCL03. Z kolei rekord bazy danych to ustrukturyzowany zestaw informacji w postaci wiersza tabeli. Powróćmy do wymogów technicznych programu. Po pierwsze, poszczególne przypadki muszą być umieszczone w osobnych wierszach, czyli stanowić osobne rekordy. Po drugie, nazwy przypadków i rekordów bazy danych muszą być jednakowe i muszą być umieszczone w pierwszej kolumnie – chodzi o to, aby program był w stanie zlokalizować odpowiadające sobie elementy. Po trzecie, w pierwszej komórce bazy (w arkuszach kalkulacyjnych jest to na ogół komórka A1) należy wpisać unikatową nazwę, jaką chcemy

wykorzystać do opisu danego zespołu atrybutów. Jest to o tyle istotne, że do jednego zestawu przypadków można przypisywać wiele klasyfikacji bez konieczności łączenia ich w jedno – ułatwia to pracę, gdy wykorzystuje się dane z różnych źródeł. Spełnienie owych warunków umożliwi sprawne połączenie bazy danych z przypadkami.

Procedura importu bazy danych sprowadza się do uruchomienia kreatora dostępnego w menu *External Data*. Etap pierwszy to wskazanie pliku do zaimportowania. Drugi to wskazanie, czy importowana klasyfikacja ma dotyczyć źródeł, czy przypadków (wybór zależy od tego, jak uporządkowany został zbiór danych: czy program ma połączyć bazę z zawartością katalogu *Sources*, czy też z określoną grupą kodów – *Cases*). Opcje do wyboru dotyczą:

- **utworzenia** w zbiorze materiałów nowych atrybutów, jeśli takowe jeszcze nie istnieją (w przypadku pierwszego importu – trzeba zaznaczyć),
- **aktualizacji** importowanej bazy o przypadki lub dokumenty nieistniejące w bazie (jeśli materiał jest uporządkowany, to nie ma to sensu);
- **zamiany wartości atrybutów dokumentów** w zbiorze o te zawarte w importowanej bazie (użyteczne w przypadku aktualizacji lub poprawek).

Etap trzeci to wskazanie lokalizacji, w której znajdują się przypadki lub źródła mające zostać połączone z rekordami bazy danych. Jeśli baza została skonstruowana poprawnie, to najwygodniej jest użyć opcji *As names*, wskazujemy wówczas jedynie katalog zawierający obiekty, a kluczem łączącym będą jednakowe nazwy przypadków/źródeł oraz rekordów importowanej bazy. Etap czwarty to wprowadzenie informacji szczególnych, takich jak oznaczenia występujące w bazie danych, odnoszące się do braków danych lub informacji „nie dotyczy” oraz – jeśli występują – formaty informacji liczbowych. Zaimportowaną bazę danych w widoku arkusza, schematu klasyfikacyjnego przedstawiają kolejne ilustracje (rys. 9 oraz 10).

Rys. 9. Zaimportowana do NVivo baza danych dotycząca liderów

Name	A: Q1A.1	B: Q1A.2	C: Q1A.3	D: Q1A.4
1: BKL01	nie	tak	tak	nie
2: BKL02	nie	nie	tak	tak
3: DCL01	tak	nie	tak	tak
4: DCL02	tak	tak	tak	tak
5: DCL03	nie	nie	nie	nie
6: JML01	nie	tak	nie	tak
7: JML02	nie	tak	tak	nie
8: JML03	tak	tak	tak	nie
9: JML04	nie	nie	nie	tak
10: JML05	nie	nie	nie	tak

Źródło: opracowanie własne.

Podgląd atrybutów pojedynczych przypadków jest możliwy dzięki wywołaniu funkcji właściwości, dostępnej po kliknięciu prawym przyciskiem myszy i wybraniu opcji *Node Properties* z menu kontekstowego.

Zakładka *Attribute Values* zawiera szczegółowy podgląd. W przypadku przypisywania atrybutów dokumentom w katalogu *Sources*, okno właściwości zostanie wywołane po kliknięciu w dokument źródłowy.

Rys. 10. Podgląd atrybutów przypadku BKL01 – w widoku właściwości

Attribute	Value
Q1A.1	nie
Q1A.2	tak
Q1A.3	tak
Q1A.4	nie

Źródło: opracowanie własne.

Podsumujmy procedurę budowania bazy danych. W celu uniknięcia żmudnej lektury (przypomnę – blisko 80 wywiadów) i kategoryzowania wypowiedzi, na co podczas badania nie było czasu, pracę zorganizowano tak, aby rozłożyć ją pomiędzy członków zespołu realizującego wywiady. Przygotowania skupiły się zatem na skonstruowaniu odpowiedniego nośnika umożliwiającego scalenie pracy wielu osób oraz pozwalającego segmentować dane. W praktyce oznaczało to stworzenie materiałów dopasowanych pod względem formalnym do możliwości programu komputerowego (dyspozycje i arkusze transkrypcyjne, arkusze informacji standaryzowanych oraz szablon bazy danych). Z powyższego wynika, że „pomnażanie brakującego czasu” dokonano się głównie dzięki sprawnej organizacji oraz wykorzystaniu nowoczesnych technik przetwarzania danych. Kolejnymi elementami procedury były autokodowanie oraz import bazy da-

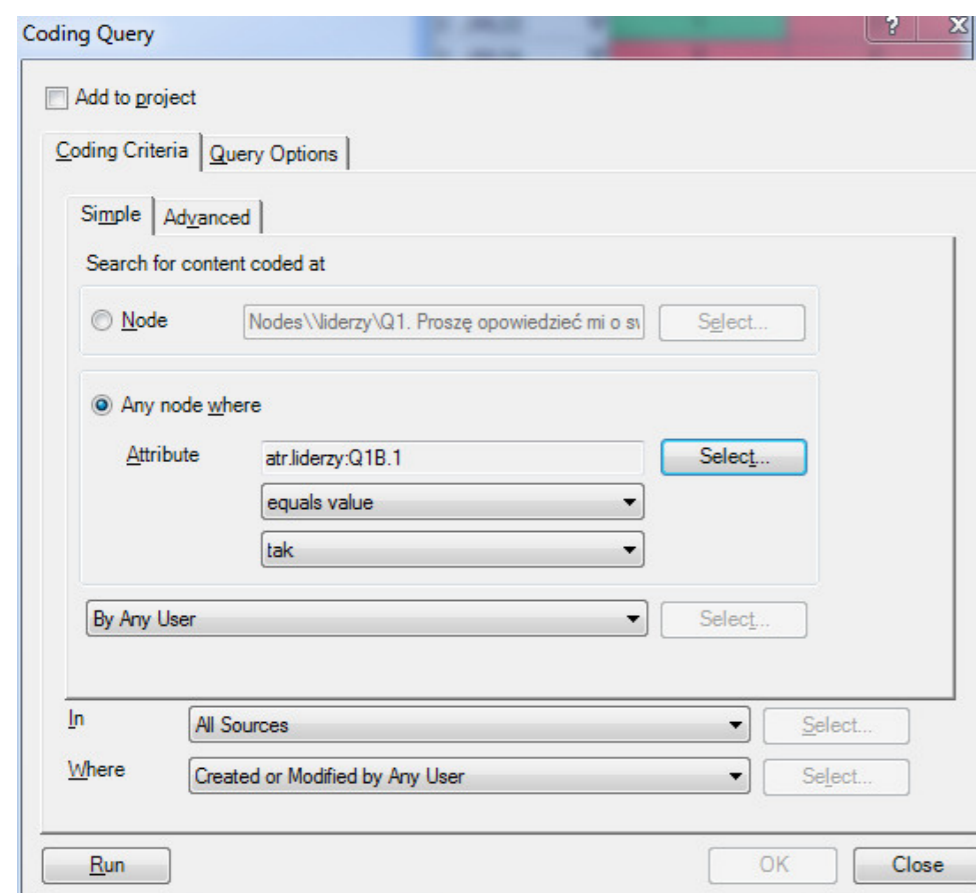
nych do programu. Od tego momentu rozpoczęła się zasadnicza praca analityka.

Usprawnienie procesu analizy danych oraz opracowywania raportu z badań polegało przede wszystkim na unikaniu czynności zbędnych. Kluczowe znaczenie miała tu organizacja pracy. W tym przypadku najważniejszym elementem była problematyka badania wraz z uzgodnionym sposobem prezentacji, odpowiadającym zamawiającemu. Założenia te powinny być dyskutowane ze zleceniodawcą, inwestorem, partnerem czy użytkownikiem danych już przy pierwszym spotkaniu inicjującym przedsięwzięcie badawcze. Potrzeby podmiotu zamawiającego badanie oraz wyobrażenia członków zespołu realizującego mogą być rozbieżne i wynikać z odmiennych oczekiwań odnośnie działalności badawczej i jej rezultatów. Zagadnienia wyrażone pytaniami w rodzaju: „Do czego to ma służyć?”, „Kto

ma z tego korzystać?” i „Ile mamy czasu?” powinny paść na początku współpracy. Tak było i w przypadku badania będącego podstawą prezentowanych tu rozwiązań. Dokumenty, które powstały po pierwszym „zebraniu”, to zarys problematyki, raportu oraz wstępna lista zagadnień do uwzględnienia w dyspozycjach. Zatem, kiedy baza danych była już kompletna, analityk wiedział, jakie zabiegi segmentacyjne należy przeprowadzić, aby uniknąć lektury i kodowania treści, które nie zawierają potrzebnych informacji. W tym miejscu czytelna jest również rola, jaką odegrali opracowujący wywiady: w badaniu przewidziano dwa szkolenia przygotowujące do prac terenowych.

Podstawowym narzędziem wykorzystanym w fazie opracowywania wyników były kwerendy macierzowe (*Matrix Coding*). Narzędzie to pozwala dokonywać przeglądu danych pod kątem występowania wypowiedzi spełniających założone warunki. W tym przypadku zapytania warunkowe – pozwalające wyszukiwać spełniające je elementy – obejmowały określoną kategorię kodową (choć trafniejszym określeniem byłaby tu kategoria tematyczna) oraz odpowiedni zestaw atrybutów. Macierz obejmująca te dwa wymiary pozwalała wybrać treści do lektury, które należało uwzględnić w raporcie. Dla przykładu przyjrzyjmy się zagadnieniu „charakterystyki prowadzonej działalności” (Q1B), poruszanemu

Rys. 11. Kwerenda kodowania z uwzględnieniem pojedynczej wartości atrybutu



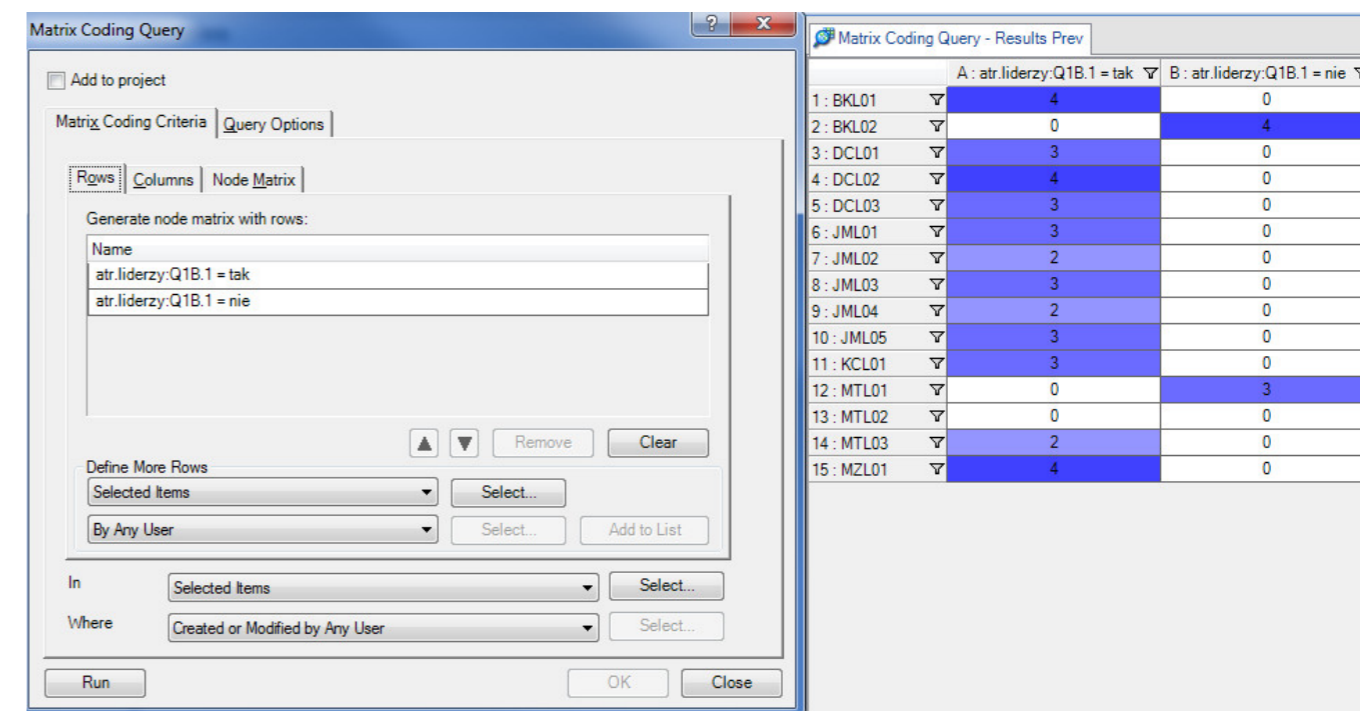
Źródło: opracowanie własne.

w rozmowach z liderem (patrz rys. 1). Ze względu na charakter projektu, w ramach którego realizowano badanie, ważne było to, aby uwzględnić fakt – przykładowo – czy badany faktycznie był wolontariuszem lub – czy brał udział w przedsięwzięciach adresowanych do osób starszych. Zastosowanie kwerendy macierzowej umożliwiło wybranie wypowiedzi tylko tych badanych, którzy – charakteryzując swoją działalność – opowiadali o przyjmowanej roli i jednocześnie faktycznie byli wolontariuszami. Opisany zabieg możliwy jest do wykonania przy użyciu zwykłego zapytania kodowego (*Coding Query*) (patrz rys. 11), ale w odróżnieniu od kwerendy macierzowej nie pozwala na podgląd rozkładu wypowiedzi w danej grupie przypadków. Kolejne ilustracja (rys. 12) przedstawia wynik zapytania, tabela zawiera w wierszach wybrane przypadki, w kolumnach – wartości atrybutów (tutaj: „tak” lub „nie”), natomiast w poszczegól-

nych komórkach widnieje liczba fragmentów wypowiedzi spełniających warunki.

Jak widać na ilustracji poniżej, kwerenda umożliwia wyeliminowanie dwóch przypadków, które nie spełniają założonego warunku (BKL02 oraz MTL01). Pozostałe wypowiedzi powinny już zostać dokładnie przeczytane i sklasyfikowane. Od decyzji badacza zależy, jaki model porządkowania zastosuje: czy będzie sprowadzał się on do prostej analizy tematycznej, czy też bardziej złożonych form kodowania (Charmaz 2009). Zamieszczona poniżej ilustracja nie jest najlepszym przykładem skuteczności segmentowania danych, bowiem tylko dwa wywiady można było wyeliminować. Pokazuje jednak, że wkomponowanie do procesu przetwarzania danych elementów automatyzacji nie musi oznaczać całkowitego zaniechania wnikliwej lektury zgromadzonego materiału.

Rys. 12. Kwerenda macierzowa z uwzględnieniem wszystkich wartości atrybutu



Źródło: opracowanie własne.

Procedura formułowania zapytania macierzowego była standardową operacją umożliwiającą dostęp do danych pozwalających nasycić informacją poszczególne obszary problematyki badawczej. Identyfikowane w obu zbiorowościach – przedstawiciele i liderów – kategorie kodowe były porównywane między sobą w celu wskazania obszarów stycznych i rozłącznych. Taki przegląd danych umożliwił opracowanie kompletnego raportu z badań. Nie byłoby to jednak możliwe, gdyby nie zastosowanie opisanej w artykule organizacji pracy opartej na możliwości przetwarzania danych w programie typu CAQDA.

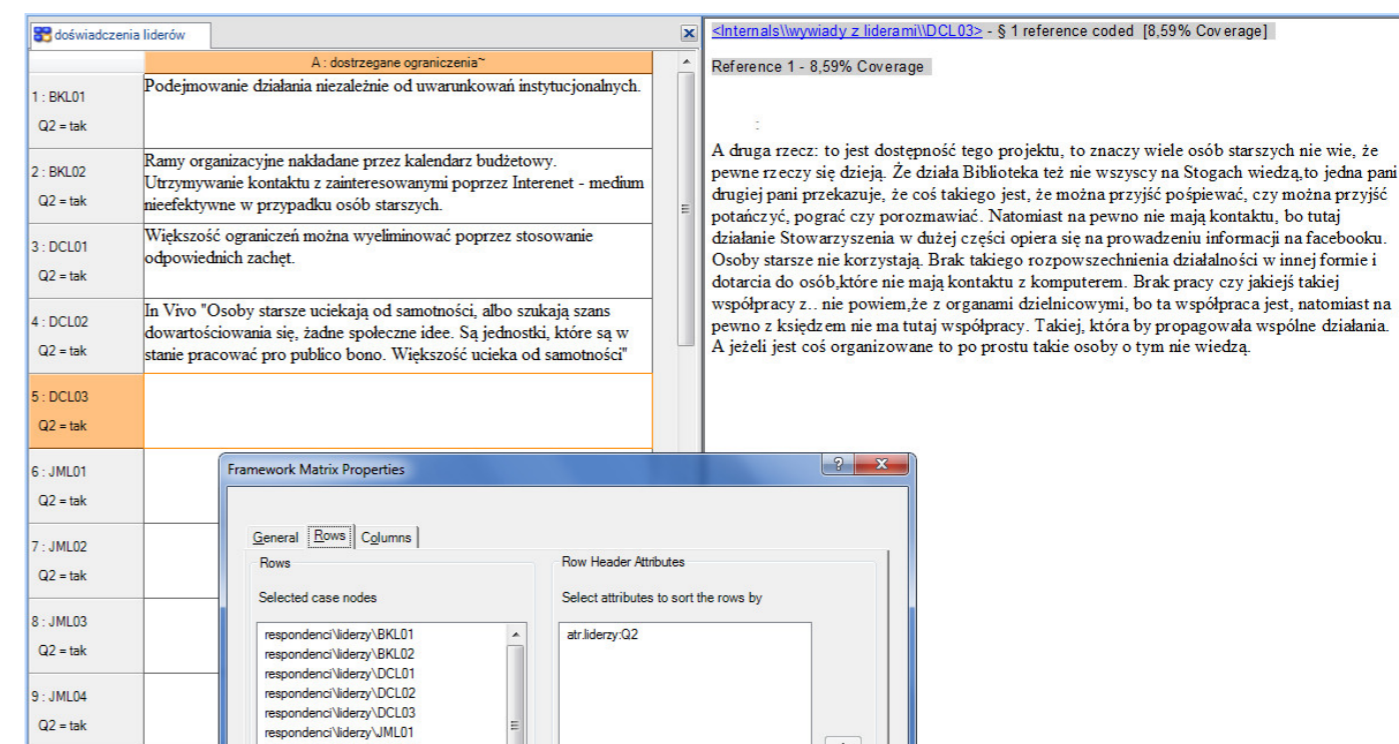
Dodatkowym narzędziem wspomagającym interpretację materiału były macierze ramowe (*Framework Matrices*). Narzędzie to ułatwia formułowanie wniosków i podsumowań, bazując na bezpośrednich odniesieniach do fragmentów materiału empirycznego spełniającego określone kryteria. Dzięki temu narzędziu można w prosty sposób odwoływać się do analizowanych porcji tekstu, co jest szczególnie istotne przy porównywaniu przypadków. Każde spostrzeżenie może zostać zanotowane bez potrzeby wywoływania dodatkowych okien dialogowych. Budowanie macierzy przypomina formułowanie kwerendy macierzowej. Uzyskany rezultat różni się jednak od zaprezentowanego na powyższej ilustracji (rys. 12).

Zasada użycia funkcji *Matrix Framework* wiąże się z łatwym dostępem do fragmentu tekstu spełniającego wybrane kryteria, na przykład zakodowanego jako „dostrzegane ograniczenia” oraz posiadającego atrybut „tak” (rys. 13). Kliknięcie dowolnej komórki natychmiast „przechwytuje” fragment

spełniający warunki, a w samej komórce można wpisywać spostrzeżenia. W macierzy można umieścić dowolną liczbę kategorii kodowych i atrybutów – jedyną barierą jest czytelność i przejrzystość. Praca z „ramkami” była szczególnie przydatna w końcowej fazie badania, kiedy ostatecznie wyselekcjonowane zbiory danych poddawane były analizie porównawczej. Wówczas w jednej macierzy znajdowały się obok siebie różnorodnie klasyfikowane treści, a nawigacja między nimi sprowadzała się do pojedynczych kliknięć, które przywoływały pożądaną treść w oknie podglądu – widocznym z prawej strony ilustracji.

Możliwości NVivo są oczywiście o wiele większe niż zaprezentowana tu konfiguracja operacji zarządzania zbiorem danych. Program jest wyposażony w bardzo bogaty moduł eksploracyjny, w którym zaimplementowano wiele narzędzi charakterystycznych dla podejść badawczych określanych jako *Text Mining* i *Data Mining*. Narzędzia takie jak klastrowanie, analiza frekwencyjna słów, *chmury tagów* czy też statystyczne algorytmy związków między źródłami i kategoriami kodowymi nie wydawały się odpowiednie do analizy indywidualnych wywiadów pogłębionych, przeprowadzonych na potrzeby pojedynczego projektu. Zbudowanie odpowiednich reguł słownikowych i zautomatyzowanych kluczy kategoryzacyjnych zajęłoby tyle czasu, ile lektura i opracowanie wyselekcjonowanych fragmentów. Gdyby materiał obejmował *korpus* tekstowy liczący kilkaset pozycji lub też realizowany projekt stanowił początek długofalowego przedsięwzięcia badawczego, to zastosowanie bardziej wyrafinowanych narzędzi automatyzujących stałoby się koniecznością i nade wszystko byłoby zasadne.

Rys. 13. Macierz ramowa



Źródło: opracowanie własne.

Przeprowadzonym w badaniu operacjom przyswiecała także idea unikania zjawiska „czarnej skrzynki”, czyli ryzyka wystąpienia operacji niekontrolowanych. Program NVivo umożliwia oparcie całego badania na następującym schemacie: autokodowanie, analiza frekwencyjna słów w poszczególnych kategoriach kodowych, klastrowanie bazujące na współczynnikach Pearsona lub Jaccarda, wspomagane dodatkowo podglądem kontekstu słów kluczowych oraz interpretacja *chmury tagów*. Byłoby to jednak zupełnie inne badanie od założonego we wniosku projektowym. Strategia taka przybliżałaby przedsięwzięcie do podejścia charakterystycznego dla analiz dyskursywnych, gdzie przedmiotem badawczego wglądu jest „użycie słów”, podczas gdy intencją przeprowadzonego badania było zdiagnozowanie określonego stanu rzeczy obejmującego nie tylko zachowania językowe.

W uniknięciu obrania niekontrolowanej „drogi na skróty” pomogły precyzyjne uzgodnienia z podmiotem zlecającym badanie oraz dobra znajomość użytkowanego narzędzia CAQDA. Znajomość specyfiki programu pozwoliła tak zaprojektować procedurę badawczą, aby wymagania techniczne NVivo nie wpłynęły na modyfikację istoty badania terenowego. Pomimo wykorzystania specyficznych arkuszy, indywidualny wywiad pogłębiony pozostał wywiadem. Pomimo zastosowania autokodowania oraz zabiegów segmentacyjnych, analiza materiału opierała się na lekturze wypowiedzi, identyfikowaniu kategorii oraz ich porównywaniu. Program CAQDA pozostał użytecznym narzędziem w ręku badacza, który musiał „skracać drogę”, ale szczęśliwie „trafił do celu”.

Bibliografia

Bieliński Jacek, Iwińska Katarzyna, Rosińska-Kordasiewicz Anna (2007) *Analiza danych jakościowych przy użyciu programów komputerowych*. „ASK”, nr 16, s. 89–114.

Błędowski Piotr, Mossakowska Małgorzata, Więcek Andrzej (2012) *Aspekty medyczne, psychologiczne, socjologiczne i ekonomiczne starzenia się ludzi w Polsce*. Poznań: Termedia Wydawnictwo Medyczne.

Brosz Maciej (2012) *Komputerowe wspomaganie badań jakościowych. Zastosowanie pakietu NVivo w analizie materiałów nieustrukturyzowanych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 98–125 [dostęp 15 grudnia 2013 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakoosciowej.org>>.

Charmaz Kathy (2009) *Teoria ugruntowana. Praktyczny przewodnik po analizie jakościowej*. Przełożyła Barbara Komorowska. Warszawa: Wydawnictwo Naukowe PWN.

Czekanowski Piotr (2012) *Spoleczne aspekty starzenia się ludności w Polsce. Perspektywa socjologii starości*. Gdańsk: Wydawnictwo UG.

Czekanowski Piotr, Załęcki Jarosław, Brosz Maciej (2013) *Gdańska starość. Portret socjologiczny mieszkańców Gdańska w wieku 65+*. Gdańsk: Fundacja Terytoria Książki.

Kiwi (2013) *Informacje o projekcie „Gdański Model Wolontariatu Seniorów – krok 1”* [dostęp 14 grudnia 2013 r.]. Dostępny w Internecie: <<http://kiwi.org.pl/event/details/666>>.

Niedbalski Jakub (2013) *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo UŁ.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 126–165 [dostęp 15 grudnia 2013 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakoosciowej.org>>.

Nowak Stefan (2007) *Metodologia nauk społecznych*. Warszawa: Wydawnictwo Naukowe PWN.

Saillard Elif Kus (2011) *Systematic Versus Interpretive Analysis with Two CAQDAS Packages: NVivo and MAXQDA*. „Forum: Qualitative Social Research”, vol. 12, no. 1 [dostęp 30 września 2011 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1518>>.

Cytowanie

Brosz Maciej (2014) *Badanie jakościowe w warunkach niedoszacowania czasu – czyli o konieczności „podążania na skróty”*. Organizacja procesu badawczego pod kątem zastosowania komputerowego wsparcia analizy danych jakościowych. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 40–58 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakoosciowej.org>.

Qualitative Research in the Condition of Underestimated Time – In Need of “Shortcuts.” The Organization of Research Process Using the Computer-Aided Qualitative Analysis

Abstract: This paper concerns the organization of research process using NVivo software. Described researches were conducted in the condition of underestimated time. The use of CAQDA (computer-aided qualitative data analysis) and proper research management enabled fitting into the timeframe imposed by the contracting institution. The work distribution and its management, preparation of properly formatted documents, and complementation of qualitative database with the collection of attributes based on standardized data significantly improved the elaboration of results.

Keywords: CAQDAS, NVivo, auto coding, data segmentation, import classification sheet data, matrix coding query, framework matrix

Jakub Niedbalski
Uniwersytet Łódzki

DOI: <https://doi.org/10.18778/1733-8069.10.2.04>

Zastosowanie oprogramowania Atlas.ti i NVivo w realizacji badań opartych na metodologii teorii ugruntowanej

Abstrakt Artykuł porusza problematykę zastosowania specjalistycznego oprogramowania wspomagającego analizę danych jakościowych w badaniach opartych na procedurach metodologii teorii ugruntowanej (MTU). Celem artykułu jest pokazanie, jakie związki istnieją pomiędzy procedurami metodologii teorii ugruntowanej a dwoma popularnymi programami z rodziny CAQDAS: NVivo oraz Atlas.ti. Artykuł ma wskazać, w jaki sposób można wykorzystać funkcje dostępne w obu programach do prowadzenia analizy opartej na MTU. W artykule pokazano zarówno techniczne, jak i aplikacyjne możliwości oprogramowania NVivo i Atlas.ti. Wskazano także na stopień zgodności rozwiązań technicznych stosowanych w obu programach do wymogów metodologii teorii ugruntowanej oraz na pewne ograniczenia i bariery, które napotkać może badacz wykorzystujących dany program komputerowy w badaniach opartych na MTU.

Słowa kluczowe komputerowe wspomaganie analizy danych jakościowych, nowe technologie, metody badań społecznych, metodologia teorii ugruntowanej, NVivo, Atlas.ti

Jakub Niedbalski, doktor socjologii, adiunkt w Katedrze Socjologii Organizacji i Zarządzania Instytutu Socjologii Uniwersytetu Łódzkiego. Specjalizuje się w komputerowej analizie danych jakościowych, metodach badań jakościowych, zagadnieniach socjologii niepełnosprawności i socjologii kultury fizycznej. Prowadzi badania poświęcone aktywizacji społecznej i fizycznej osób z niepełnosprawnością. Autor książek: *Życie i praca w domu pomocy społecznej. Socjologiczne studium interakcji personelu z upośledzonymi umysłowo podopiecznymi* (Wydawnictwo UŁ, Łódź 2013), *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe*

wspomagające analizę danych jakościowych (Wydawnictwo UŁ, Łódź 2013) oraz *Komputerowe wspomaganie analizy danych jakościowych. Zastosowanie oprogramowania NVivo i Atlas.ti w projektach badawczych opartych na metodologii teorii ugruntowanej* (Wydawnictwo UŁ, Łódź 2014).

Adres kontaktowy:

Katedra Socjologii Organizacji i Zarządzania
Instytut Socjologii Uniwersytetu Łódzkiego
ul. Rewolucji 1905 r. 41, 90-214 Łódź
e-mail: jakub.niedbalski@gmail.com

CAQDAS to inaczej komputerowe oprogramowanie wspomagające analizę danych jakościowych. Termin ten został wprowadzony przez naukowców z University of Surrey w Wielkiej Brytanii na przełomie lat osiemdziesiątych i dziewięćdziesiątych ubiegłego wieku i odnosi się do kategorii oprogramowania służącego do realizacji projektów badawczych prowadzonych zgodnie ze zróżnicowanymi podejściami analitycznymi z zakresu badań jakościowych (Seale 2008: 232–233). Obecnie możemy obserwować niezwykle dynamiczny rozwój tego rodzaju oprogramowania, a lista dostępnych programów z każdym rokiem staje się coraz dłuższa (zob. Lewins, Silver 2004). Obok narzędzi stosunkowo prostych i o ograniczonych możliwościach mamy narzędzia tak rozbudowane, jak Atlas.ti czy NVivo, które udostępniają badaczowi opcje tworzenia powiązań między kodami, zaawansowanego przeszukiwania danych, generowania hipotez, a w dalszej kolejności także konstruowania teorii (Kelle 2005: 486; Fielding 2007: 463). Wskazane oprogramowanie pomaga również w tworzeniu rozbudowanych zbiorów danych i ich wszechstronnemu porządkowaniu. Z tego względu może być stosowane przez badaczy reprezentujących różne szkoły teoretyczne i podejścia metodologiczne. Zdaniem Gibbsa (2011: 20), niektórzy mogą głównie skoncentrować się na zarządzaniu danymi, wykorzystując takie funkcje, jak grupowanie, gromadzenie, segregowanie czy proste przeszukiwanie informacji. Z kolei inni będą chcieli za ich pomocą przeprowadzić cały proces analizy, łącznie z wygenerowaniem teorii i sformułowaniem końcowych wniosków. Dlatego Atlas.ti oraz NVivo uważa się za narzędzia

wspomagające projekty badawcze z zakresu analizy dyskursu, metodologii teorii ugruntowanej, analizy konwersacyjnej, etnografii, badań opartych na fenomenologii, a także innych, w tym mieszanych metod badawczych (Schönfelder 2011). Z tego względu grupa badaczy, do których adresowane są oba programy, jest dość liczna, a fakt wsparcia w postaci konsultacji, doradztwa oraz szkoleń zapewnianych przez producentów tych pakietów sprawia, że nawet niedoświadczeni użytkownicy mogą w miarę szybko nabyć podstawowe umiejętności pozwalające na wykorzystywanie programów w prowadzonych przez siebie projektach badawczych. Warto też dodać, że NVivo i Atlas.ti to narzędzia stale udoskonalane, nad którymi trwają nieprzerwane i intensywne prace polegające między innymi na modyfikacji istniejących, ale także dodawaniu nowych funkcji (Gibbs 2011; zob. też Brosz 2012; Niedbalski, Ślęzak 2012).

Z drugiej strony w 2010 roku Michael Jones i Kieren Diment (2010: 10), dokonując przeglądu publikacji naukowych, ustalili, że najpowszechniej stosowanymi narzędziami informatycznymi z rodziny CAQDAS są NVivo oraz Atlas.ti. Ci sami autorzy stwierdzili, że przeważająca liczba badaczy, którzy posługują się tymi programami, wykorzystuje w swojej pracy metodologię teorii ugruntowanej (por. Budziszewska 2010: 89–94).

Biorąc powyższe pod uwagę, podjęta została decyzja o przedstawieniu takich narzędzi CAQDAS, które w najlepszy sposób spełniają rolę wspomagającą w działaniach analityka kierującego się zasadami MTU (Gibbs 2011: 20).

Badacz jako użytkownik oprogramowania NVivo oraz Atlas.ti

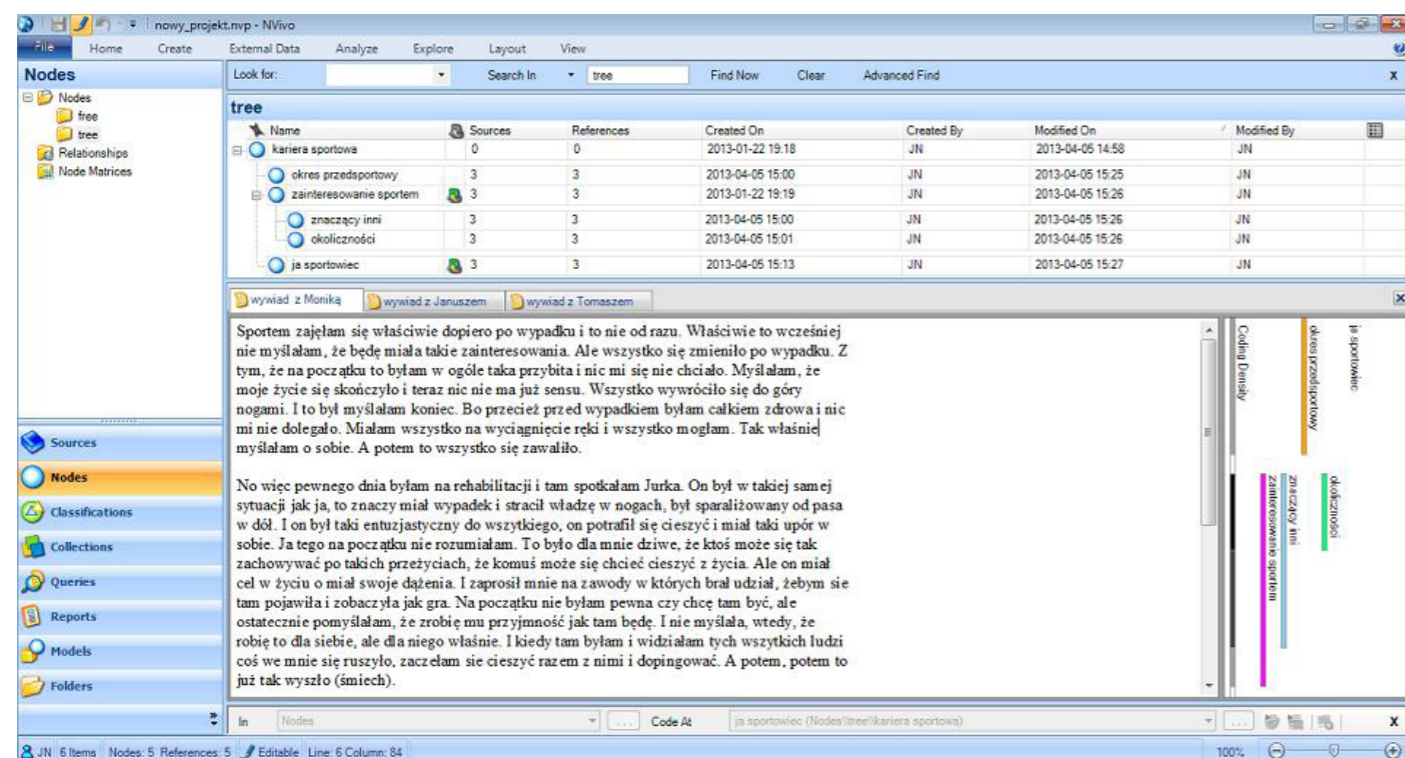
Rozważania nad wykorzystywaniem obu programów w kontekście realizacji projektów badawczych opartych na metodologii teorii ugruntowanej warto jest rozpocząć od charakterystyki cech użytkowych NVivo i Atlas.ti, przede wszystkim zaś opisu interfejsu, rozlokowania poszczególnych funkcji, a także porównania stosowanego w nich nazewnictwa opcji oraz ogólnej nomenklatury, jaką wprowadzili projektanci jednego oraz drugiego oprogramowania (Niedbalski 2014: 24).

Zarówno NVivo, jak i Atlas.ti jest swoistym środowiskiem, w którym badacz pracuje i wykonuje określone czynności zgodnie z tak zwaną „archi-

tekturą oprogramowania”, a więc technicznymi rozwiązaniami zaimplementowanymi przez jego konstruktorów (Saillard 2011; Niedbalski, Ślęzak 2012: 128). Owa organizacyjna struktura oprogramowania ma swoje odzwierciedlenie w układzie interfejsu obydwu programów.

W NVivo główny interfejs programu jest podzielony na kilka obszarów: na górze menu główne w formie zakładek z opcjami do wyboru, posegregowanymi według grup poszczególnych narzędzi, po lewej stronie znajduje się menu nawigacji umożliwiające dostęp do najważniejszych funkcji wspomagających analizę danych. Najwięcej miejsca zajmuje obszar roboczy, w którym wyświetlana jest zawartość folderów, a także treść poszczególnych plików (Niedbalski 2014: 28).

Rys. 1. Okno programu NVivo 9

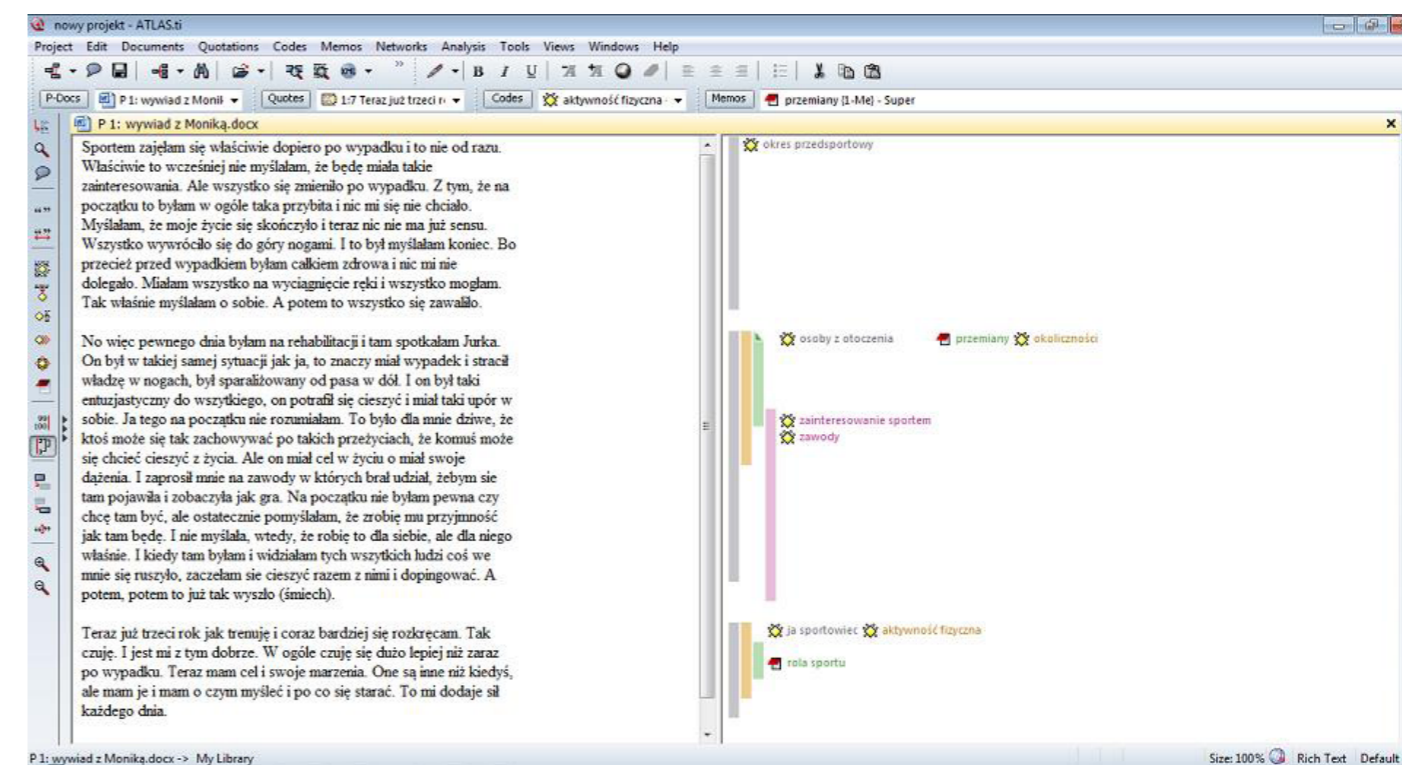


Źródło: opracowanie własne na podstawie programu NVivo 9.

Z kolei po uruchomieniu Atlas.ti i otwarciu wybranego (a wcześniej już dodanego do projektu) dokumentu pojawi się główne okno programu, w którym można wyróżnić kilka kluczowych komponentów. Patrząc od samej góry, w lewym rogu okna widoczna jest nazwa aktualnie wyświetlanego i edytowanego projektu HU. Nieco poniżej znajduje się rozwijane menu oraz paski narzędzi. Pod paskami narzędzi znajdują się natomiast cztery, ważne z punktu widzenia prowadzonej przez badacza analizy, pola z rozwijanymi listami. Są to w kolejności od lewej strony: Dokumenty Pierwotne

(Primary Documents), Cytaty (Quotations), Kody (Codes) oraz Mema (Memos). Po prawej stronie okna mamy natomiast pasek narzędzi szybkiego dostępu do najczęściej używanych opcji, takich jak: proste przeszukiwanie tekstu, edytowanie dokumentu pierwotnego, tworzenie kodów, memo, widok utworzonych modeli czy zmiana wielkości dokumentu. Zdecydowanie największą przestrzeń okna zajmuje obszar roboczy, w którym wyświetlane są elementy projektu wybrane przez użytkownika (m.in. tekst, zdjęcia, grafika, notatki itd.) (Niedbalski 2014: 26).

Rys. 2. Okno programu Atlas.ti 7



Źródło: opracowanie własne na podstawie programu Atlas.ti 7.

Warto zaznaczyć, że nomenklatura, jaką operują obydwaj programy w zakresie poszczególnych funkcji, w niektórych momentach różni się. W związku z tym warto o nich wspomnieć, zwłaszcza w kontekście procedur, jakie obecne są w ramach metodologii teorii

ugruntowanej. W sposób bardziej obrazowy prezentowane różnice w kontekście terminologii, jaką posługuje się metodologia teorii ugruntowanej, a która występuje w programie NVivo i Atlas.ti, prezentuje poniższa tabela (por. Salinger, Plonka, Prechelt 2007).

Tab. 1. Zestawienie nazewnictwa MTU oraz oprogramowania NVivo i Atlas.ti.

MTU	Atlas.ti	NVivo
phenomena (zjawiska)	quotations	references
conceptualization (noty)	memos	memos
concepts (kody)	concept/code	nodes
properties (własności)	concept/code	node classification
categories (kategorie)	families	tree nodes
relationships (relacje)	relationships/relations	relationships

Źródło: opracowanie własne (Niedbalski 2014: 32).

Ważniejsze od samej nomenklatury, z jaką spotykamy w NVivo oraz Atlas.ti, jest to, co faktycznie kryje się pod określonymi pojęciami, a także to, jak konkretne rozwiązania dotyczące struktury (czy inaczej mówiąc, architektury wewnętrznej) danego programu wpasowują się w wymogi stawiane przez metodologię teorii ugruntowanej. Jest to kluczowe w kontekście realizacji projektów badawczych, bowiem sama nomenklatura, choć czasami (zwłaszcza zaś na początku użytkowania danego programu) może być myląca i uciążliwa, to jednak użytkownik jest w stanie najszybciej się z nią oswoić, w przeciwieństwie do sposobu wykorzystania poszczególnych funkcji oferowanych przez oprogramowanie, które nie tylko należy poznać, ale też umieć w odpowiedni sposób zaadaptować do wymagań określonej metody i strategii badawczej.

Badacz jako analityk

Oba opisywane programy to narzędzia niezwykle rozbudowane, oferujące szereg funkcji pozwalających na wykonywanie zaawansowanych analiz jakościowych, począwszy od prostego grupowania i zliczania danych po tworzenie hipotez i generowanie teorii. Oba

też uważane są za narzędzia wspomagające projekty badawcze z zakresu analizy dyskursu, metodologii teorii ugruntowanej, analizy konwersacyjnej, etnografii, badań opartych na fenomenologii, a także innych, w tym mieszanych metod badawczych (Schönfelder 2011; Niedbalski, Ślęzak 2012: 132). Niemniej jednak, ze względu na cel artykułu, przedstawione zostały w nim wybrane opcje oraz taki ich sposób wykorzystania, który odpowiada prowadzeniu badań zgodnie z metodologią teorii ugruntowanej.

W związku z tym musimy zwrócić uwagę na dwa główne etapy, na jakie można podzielić pracę analityka, a które odpowiadają określonym zestawom czynności wykonywanych za pomocą oprogramowania CAQDAS. Są to mianowicie **poziom wstępny** oraz **poziom konceptualny**¹. Pierwszy z nich obejmuje przede wszystkim działania związane z technicznym przygotowaniem materiałów (źródeł danych analitycznych), tworzeniem projektu, gromadze-

¹ Autorzy programu Atlas.ti wprowadzają na przykład podział pracy badacza na **poziom tekstowy** – m.in. kodowanie, segregowanie kodów, tworzenie kategorii, nanoszenie notatek – oraz **poziom konceptualny** – poszukiwanie związków i zależności między kodami/kategoriami oraz budowanie modelu.

niem danych i ich formalnym segregowaniem oraz grupowaniem. W tym zakresie mieszczą się także takie aktywności użytkownika oprogramowania, jak opisywanie danych i ich transkrypcja. Jest to zatem cały szereg czynności poprzedzających proces właściwej pracy analitycznej i konceptualnej (Niedbalski 2014: 47). Niemniej jednak jest to równie istotny etap działalności badacza jak sama analiza, bowiem wszystkie wskazane powyżej czynności stanowią grunt dla dalszej interpretacji danych i procesu teoretyzowania z budowaniem teorii włącznie².

Poziom konceptyjny dotyczy natomiast całego procesu analitycznego, który opiera się na takich czynnościach, jak kodowanie danych, ich tematyczne segregowanie i przyporządkowywanie do określonych kategorii, a następnie łączeniu i poszukiwaniu związków, a także istniejących między nimi zależności. Jest to proces nieodłącznie związany z budowaniem hipotez oraz ich weryfikowaniem w toku dalszej analizy i interpretacji danych. Pomocne w tym zakresie są – poza wspomnianym już kodowaniem i kategoryzowaniem – także pisanie not teoretycznych w postaci memo, możliwość przegrupowania i ciągłego modyfikowania wygenerowanych danych, wykorzystanie wszelkiego rodzaju narzędzi przeszukiwania używanych w procesie nie tylko kwerendy tekstu czy kodów bądź kategorii, ale też wyszukiwania danych w oparciu o pytania badawcze. Są to również narzędzia zapewniające możliwość wizualizacji wyników procesu analizy danych oraz tworzenia modeli odwzorowujących sieć powiązań i zależności, jakie istnieją pomiędzy elementami projektu (Niedbalski 2014: 269).

² Warto jednak podkreślić, że proces przygotowawczy, w tym transkrypcja danych, może być, w zależności od przyjętej metodologii badawczej, częścią samej analizy danych.

Działania te zgodne są z kolei z zasadami metodologii teorii ugruntowanej, bowiem celem badacza jest stopniowe przechodzenie od materiału empirycznego na coraz wyższe poziomy abstrakcyjnego rozumowania poprzez tworzenie zróżnicowanych hierarchicznie kategorii i ich własności – aż do konstruowania hipotez i teorii. Logika procesu badawczego opiera się na dążeniu do coraz wyższego poziomu konceptualnego, a co za tym idzie – oderwania się od danych ku teoretyzowaniu. Kluczową rolę odgrywa w tym kontekście proces kodowania, a więc przypisywania partiom materiału określonych etykiet odzwierciedlających ich sens i znaczenie nadawane im przez aktorów społecznych i odwzorowane przez badacza. Działaniom tym towarzyszą określone procedury metodologiczne poprawności, do których w przypadku MTU należą między innymi teoretyczne pobieranie próbek (*theoretical sampling*), procedura ciągłego porównywania (*constant comparative method*), kodowania, pisanie not.

W związku z tym w dalszej części artykułu opisane zostaną najważniejsze ze wspomnianych powyżej procedur metodologii teorii ugruntowanej w kontekście możliwości ich zastosowania w analizach prowadzonych przy wsparciu oprogramowania komputerowego.

Poziom przygotowawczy – edycja, segregowanie, opisywanie i transkrypcja danych

Generalną zasadą w przypadku wszystkich programów CAQDA jest to, że każdy z nich stanowi przede wszystkim narzędzie służące do tworzenia bazy danych, które następnie – przy użyciu

zaimplementowanych w nim funkcji – można w różny sposób przetwarzać. Podobnie jest w przypadku programów NVivo oraz Atlas.ti, które pozwalają na tworzenie rozbudowanej bazy danych. W programie NVivo nosi ona nazwę *Project* (Projekt). W przypadku Atlas.ti autorzy programu wprowadzili na określenie bazy danych pojęcie *Hermeneutic Unit* (Jednostka Hermeneutyczna). Samą bazę można określić jako najważniejszą strukturę danych grupującą całościowy materiał empiryczny z zakresu określonej problematyki badawczej. W związku z tym baza danych zazwyczaj przyjmuje nazwę od tematyki prowadzonych badań (Niedbalski 2014: 49).

Ważną cechą obu programów jest ich uniwersalność objawiająca się tym, że oferują one możliwość opracowywania materiałów bardzo różnego rodzaju. Oprócz tekstu są to także zdjęcia czy dane audio oraz wideo. Co więcej, oba programy obsługują nie tylko różne rodzaje danych, ale też wiele ich formatów, co sprawia, że ich zakres zastosowania niewątpliwie wzrasta.

Inną ważną cechą programów NVivo oraz Atlas.ti jest to, że pozwalają one na edytowanie danych i dokonywanie modyfikacji różnych elementów projektu. Oba programy umożliwiają dokonywanie modyfikacji i poprawek w obrębie danych tekstowych. Dotyczyć one mogą między innymi zmiany czcionki, jej koloru czy użycia innych opcji formatowania, znanych z popularnych edytorów tekstowych. Nie są to być może bardzo wyszukane operacje, jednak z całą pewnością powinny wspomóc badacza w celu, do jakiego zostały powołane, a więc oznaczania określonych fragmentów tekstu

czy wyróżniania ich tak, aby pogrupować je zgodnie z ich wartością analityczną.

Ponadto NVivo oraz Atlas.ti pozwalają na swego rodzaju kompilację danych źródłowych, polegającą na włączaniu (wklejaniu) wybranych elementów w istniejące dokumenty tekstowe. Mogą to być zarówno fragmenty innych tekstów, jak i tabel, wykresy oraz materiały wizualne w postaci zdjęć.

Ważną cechą programów NVivo oraz Atlas.ti jest również to, że pozwalają one na porządkowanie różnych elementów projektu (zarówno materiałów źródłowych, jak i wszelkich informacji będących wytworem badacza zaangażowanego w analizę danych – a więc *de facto* wytworów analizy). Jest to możliwe między innymi dzięki grupowaniu danych (źródeł) w osobne foldery, zgodnie z preferencjami badacza. Jedną z takich sytuacji jest możliwość jaką daje program NVivo, wprowadzając system porządkowania danych zamieszczanych w katalogu źródeł wewnętrznych (*Internals*) bądź źródeł zewnętrznych (*Externals*). Ponadto program NVivo pozwala na tworzenie zgrupowań danych (*Collections/Sets*), zaś Atlas.ti na tworzenie tak zwanych rodzin dokumentów (*Family*).

Funkcją dostępną w obydwu programach, która pełni rolę przygotowawczą do właściwej analizy danych, jest opcja transkrypcji, czyli przełożenia w możliwie najwierniejszy sposób nagrań audio bądź wideo na formę tekstową. W Atlas.ti wprowadzono możliwość utworzenia bezpośrednio w programie transkrypcji danych, ale także synchronicznego łączenia zapisu z pierwotnym nagraniem. Oznacza to, że można czytać taki zapis,

a jednocześnie odsłuchiwać odpowiednie fragmenty pików audio/wideo i w ten sposób nie tracić takich informacji, jak ton głosu, sposób wypowiedzenia się i tak dalej, które mogą być niezwykle istotne w procesie interpretacji danych. Użytkownik w tekście transkrypcji, jak i na osi czasu pliku audio/wideo, zaznacza dokładne punkty, które odpowiadają zapisowi nagrania. W ten sposób precyzyjnie, wygodnie i szybko można odsłuchiwać wybrane fragmenty i równocześnie czytać ich transkrypcje (Niedbalski 2014: 63–71).

Tak jak Atlas.ti, także program NVivo posiada wbudowane narzędzia, które służą do tworzenia transkrypcji dla danych audio i wideo. Zapis taki, który *de facto* stanowi opis zawartości nagrania, może być następnie pomocy, gdy badacz chce użyć innych opcji programu, takich jak wyszukiwanie zawartości danych źródłowych czy określanie frekwencji występowania określonych informacji (bądź słów) w materiałach źródłowych.

W kontekście metodologii teorii ugruntowanej, opisane powyżej możliwości NVivo oraz Atlas.ti, które dotyczą tworzenia obszernej bazy danych, a także nieustannego ich dodawania i edytowania, mają istotne znaczenie w perspektywie realizacji procedury wysycania kategorii (*theoretical saturation*). Dzieje się tak, ponieważ elastyczność w „uzupełnianiu” bazy sprawia, że badacz może posiłkować się danymi do momentu, w którym stwierdzi, iż żadne nowe informacje już się w nich nie pojawiają, zaś każdy kolejny przypadek potwierdza słuszność dotychczasowych wniosków. Jednocześnie możliwość wykorzystania różnych źródeł danych, jakie badacz pragnie włączyć do swojego

projektu, pozwala wprowadzić w życie postulat B. Glasera *all is data* (pod warunkiem, że dane dobierane są zgodnie z procedurą teoretycznego pobierania próbek [*theoretical sampling*] [Glaser 1978; Strauss, Corbin 1990; Glaser, Holton 2004; 2010]). Ma to też istotne znaczenie w procesie triangulacji danych.

Z kolei porządkowanie danych zapewnia badaczowi stałą kontrolę nad gromadzonym materiałem i pozwala na bieżąco śledzić występowanie podobieństw oraz różnic pomiędzy poszczególnymi przypadkami. Możliwości, jakie w zakresie gromadzenia materiałów dają oba programy, pozwalają również badaczowi panować nad znacznymi nawet ilościami danych. Natomiast opisywany proces transkrypcji oraz powiązywania fragmentów plików audio/wideo z tekstem daje możliwość nie tylko przygotowania materiałów do analizy, ale też łatwiejszy do nich dostęp, co z perspektywy MTU zdecydowanie ułatwia pracę analityka, bowiem, choć jego celem jest stopniowe odchodzenie od danych opisowych ku wyższym poziomom konceptualizacji, to jednak badacz wielokrotnie, także już po zakodowaniu informacji, będzie powracał do materiałów pierwotnych. W tym kontekście rysuje się więc istota tego rodzaju rozwiązań technicznych zastosowanych w programach, które dobrze współgrają z ideą działania badawczego opartego na metodologii teorii ugruntowanej.

Poziom koncepcyjny

W dalszej części artykułu staram się pokazać, że niektóre z rozwiązań przyjętych przez twórców NVivo oraz Atlas.ti mogą być szczególnie przydatne dla

badaczy wykorzystujących metodologię teorii ugruntowanej. Rozwiązania zaaplikowane w obydwu programach pozostają bowiem w zgodzie z procedurami MTU i w efekcie wspomagają proces budowania teorii. W związku z tym poniżej pokazuję, w jaki sposób można korzystać z określonych, wybranych narzędzi, w które wyposażone są oba programy, aby realizować badania zgodnie z wymaganiami stawianymi przez metodologię teorii ugruntowanej.

Wyodrębnienie wątków („tematyacja”)

Podstawową jednostką analizy prowadzonej zgodnie z metodologią teorii ugruntowanej są fragmenty danych, które mogą przybrać postać „porcji” informacji, wydzielonych ze względu na interesujące badacza kwestie czy szerzej poruszane przez rozmówcę tematy. Strukturalnie mogą to być pojedyncze zwroty, wersy, zdania lub akapity, którym badacz może przypisać określone kody.

Oba programy, Atlas.ti i NVivo, wyposażono w narzędzia umożliwiające wydzielanie, edytowanie i opisywanie takich fragmentów danych. W programie Atlas.ti podstawową jednostką analizy są cytaty (*quotations*), zaś w NVivo odniesienia (*references*). W obydwu przypadkach stanowią one wybrany przez badacza segment tekstu bądź obszaru zdjęcia lub fragmentu ścieżki pliku audio/wideo, wyróżniony ze względu na jego walory poznawcze. Są to różne pod względem wielkości części materiałów źródłowych, które według badacza stanowią ważne elementy dla dalszego procesu analitycznego (Niedbalski 2014: 72).

Przy czym pomiędzy cytatami a odniesieniami istnieją pewne różnice, o których trzeba w tym

miejszu koniecznie wspomnieć. Przede wszystkim *quotations* w Atlas.ti mogą być tworzone jako niezależne elementy projektu, które następnie można powiązać z innymi składnikami bazy bądź poddać procesowi kodowania. Natomiast w przypadku *references* w NVivo, odniesienia są zawsze skojarzone z określonym kodem (bądź kodami) i zostają utworzone w momencie kodowania materiałów źródłowych.

Kodowanie i generowanie kategorii analitycznych

Twórcy CAQDAS, w tym NVivo i Atlas.ti, za jeden z głównych celów przy budowie oprogramowania postawili sobie zapewnienie możliwości wykonywania czynności kodowania. Z tej perspektywy programy NVivo oraz Atlas.ti wydają się spełniać oczekiwania badaczy wykorzystujących w swoich analizach metodologię teorii ugruntowanej, zwłaszcza że kodowanie i generowanie kategorii, a także tworzenie związków między nimi stanowią znaczną część pracy analityka posługującego się MTU. Wynikiem kategoryzacji i łączenia kodów jest konstruowanie hipotez, a na ich podstawie budowanie teorii (Niedbalski 2014: 84).

Wykonywanie kodowania zarówno w programie Atlas.ti, jak i NVivo odbywa się poprzez zaznaczanie fragment po fragmencie tekstu, któremu następnie nadaje się określone etykiety. Przy czym w procesie kodowania najważniejsze jest, aby starać się uwzględnić jak najwięcej informacji zawartych w tekście, czyniąc to jednak w sposób sensowny, sprzyjający rozwojowi koncepcji analitycznych oraz tak, aby widać było w nich dążenie badacza do teoryzowania (Stauss, Corbin 1990).

Rola kodów w programie Atlas.ti oraz NVivo jest zbieżna z funkcją, jaką posiadają one w metodologii teorii ugruntowanej. Z technicznego punktu widzenia można jednak wyodrębnić kilka zasad towarzyszących ich tworzeniu. Po pierwsze, możliwe jest tworzenie dowolnie dużej liczby kodów; po drugie, kody powinny być zbudowane z niewielkiej liczby słów; po trzecie zaś, jeśli kod jest zbyt długi, warto jego treść przenieść do komentarza/adnotacji, a samą nazwę kodu skrócić. Przede wszystkim zaś należy mieć na uwadze, że zarówno NVivo, jak i Atlas.ti umożliwiają kodowanie danych na kilka, zgodnych z MTU, sposobów wykonywania owej czynności. Po pierwsze, jest to kodowanie otwarte (*open coding*), które polega na utworzeniu nowego kodu wygenerowanego na podstawie analizy danych lub utworzeniu tak zwanego „pustego” kodu, czyli takiego, którym nie oznaczono żadnego fragmentu danych ani cytatu. Po drugie, to kodowanie in-vivo (*in-vivo coding*), czyli takie, gdzie nazwą powstałego kodu będzie fragment dokumentu pierwotnego (Gibbs 2011: 104). Po trzecie, jest to kodowanie na podstawie listy utworzonych już kodów, a więc ponowne użycie istniejącego kodu do oznaczenia nim nowego fragmentu danych bądź innego niż pierwotnie cytatu. Po czwarte, możemy również skorzystać z opcji kodowania na podstawie ostatnio używanego kodu (Niedbalski 2014: 85–97).

W obydwu programach pojawia się także piąta opcja kodowania, oparta na wynikach uprzednio przeprowadzonego przeszukiwania słownikowego danych. Narzędzie to jest używane w celu szybkiego (i zautomatyzowanego) wyszukiwania podobnych fraz czy wyrazów w tekście. Jeśli przepro-

wadzi się, zwłaszcza we wstępnej fazie, kodowanie otwarte i na tej podstawie stwierdzi, że fragmenty, które zostały zakodowane tym samym kodem, są do siebie podobne ze względu na występujące w nich słowa czy frazy (a jednocześnie badacz przeczuwa, że kody te będą istotne w dalszej analizie), można takie słowa bądź frazy wykorzystać do odszukania kolejnych fragmentów, które mogłyby być zakodowane w ten sam sposób, a więc przy użyciu tego samego kodu. Ponadto często zdarza się, że fragmenty, jakim został już przypisany jakiś kod, zawierają wyrażania, słowa bądź frazy, które mogą pojawić się również w innych partiach materiału i które sugerują podobieństwo wątków (Gibbs 2011: 219). W ten sposób można także zaktualizować, a nawet wygenerować nowy klucz kodowy. Niemniej jednak jest to sposób kodowania, który może być zawodny, bowiem występowanie pewnych słów we fragmentach danych nie oznacza od razu, że ich treść będzie faktycznie odpowiadała definicji kodu. Z drugiej strony wykorzystanie takiej procedury kodowania może spowodować, że nie zostaną zakodowane te fragmenty, które są związane z opracowywanym przez badacza wątkiem, ale w których dany rozmówca nie posługiwał się słowami stanowiącymi podstawę przeszukiwania. Ponadto wydaje się ona zbyt schematyczna, ponieważ najlepiej sprawdza się w przypadku przeszukiwania materiałów przynajmniej częściowo ustrukturyzowanych, które dla badacza MTU nie stanowią głównego źródła informacji.

Ponieważ dla analityka wykorzystującego w swojej pracy metodologię teorii ugruntowanej ważne są takie informacje, jak liczba fragmentów, które zostały zakodowane tym samym kodem, oraz to,

z iloma kodami ów kod jest powiązany, twórcy obu programów starali się przedstawić takie wiadomości w jak najbardziej przejrzysty, a jednocześnie wygodny dla badacza sposób. W Atlas.ti pierwsza cyfra znajdująca się przy kodzie informuje o tym, ile razy był on wykorzystany do oznaczenia różnych segmentów danych (*groundedness* – zakorzenie). Druga cyfra to informacja o tym, z iloma innymi kodami ów kod jest powiązany (*density* – spoistość). Obie liczby umieszczone są w nawiasie znajdującym się po nazwie kodu. W niektórych przypadkach za nawiasem może się także pojawić charakterystyczny znaczek „wężyka” (~) wskazujący na to, że taki kod posiada również własny komentarz. Z kolei w programie NVivo mamy podobne oznaczenia, choć są one podawane odwrotnie. Najpierw mamy więc informację o tym, z iloma kodami został powiązany dany kod, a następnie, ile razy został on wykorzystany do zakodowania różnych fragmentów dokumentów źródłowych. Znajdująca się zaś obok nazwy kodu ikona wskazuje na to, czy z danym kodem zostało powiązane memo (nota teoretyczna).

Zgodnie z metodologią teorii ugruntowanej, badacz stara się generować kody i kategorie na podstawie danych empirycznych. Nie oznacza to jednak, że kody raz utworzone nie mogą podlegać zmianie. Przeciwnie, według założeń MTU, kody mogą być modyfikowane, a raz nadany kod może zmienić swoją nazwę, zostać połączony z innym kodem bądź też w miarę, jak będzie postępował proces teoretyzowania, może być zastąpiony przez inny. Również takim czynnościom wychodzą na przeciw konstruktorzy obu prezentowanych tutaj programów. Oznacza to, że zarówno w NVivo, jak

i Atlas.ti możemy istniejące kody kopiować (*copy*), wycinać (*cut*), usuwać (*delete*), zmieniać ich nazwę (*rename*) czy włączać jeden kod w drugi (*merge*). Tym samym, możemy dokonywać porządkowania wygenerowanych kodów i kategorii, przechodząc na coraz wyższy poziom konceptualny. Przy czym porządkowanie kodów polega także (poza ich łączeniem, usuwaniem czy zmianą nazwy) na tworzeniu między nimi zależności, z uwzględnieniem relacji nadrzędności–podrzędności. Podobne rodzaje kodów lub kody odnoszące się do podobnych zjawisk zostają umieszczone na tym samym szczeblu w hierarchii (Gibbs 2011: 134–135). NVivo pozwala na uchwycenie takiej struktury poprzez budowanie drzewa kategorii. W samym zaś drzewie kategorii można, w miarę jak będzie postępował proces analityczny, dokonywać określonych zmian i modyfikacji. Mogą one polegać na przykład na zmianie położenia względem siebie kategorii pozostających na tym samym poziomie w hierarchii. Funkcję tę można wykorzystać w celu określenia „ważności” danej kategorii w ramach dokonywanych ustaleń analitycznych, przesuując ją powyżej lub poniżej innych kategorii. Z kolei program Atlas.ti posiada w tym zakresie następujące rozwiązania. Aby utworzyć strukturę, w której wyraźniej widoczne będą określone zależności pomiędzy kodami, należy najpierw użyć funkcji *Explorera*. Przy czym wskazane narzędzie służy do ukazania nie tylko hierarchicznej struktury kodów, lecz generalnie dostępnych elementów projektu. *Explorer* umożliwia zintegrowane i zorganizowane wyświetlanie wszystkich heterogenicznych elementów, które są tworzone w projekcie. Są to dokumenty pierwotne, cytaty, kody, notatki, rodziny i sieci (Niedbalski 2014: 100–106).

Drzewo kategorii i jego hierarchiczna struktura pozwalają badaczowi na stopniowe koncentrowanie się na wybranych gałęziach oraz poszczególnych kategoriach, a więc pełnią istotną rolę w procesie kodowania selektywnego i stopniowego kryształizowania się kategorii centralnej. Jednocześnie pozwalają na dokładniejsze określenie charakteru i zakresu relacji łączących wygenerowane kategorie, przez co sprawiają, że w procesie kodowania realne staje się stosowanie paradygmatu kodowania, a więc doprecyzowywanie kategorii, ich rozbudowywanie, odnoszenie jednych do drugich lub szukanie wzajemnych powiązań między nimi (Gibbs 2011: 98). Na paradygmat kodowania składają się zatem warunki przyczynowe wystąpienia zjawiska, warunki interweniujące, interakcje, strategie i taktyki działania, a także konteksty oraz konsekwencje działań (Strauss 1987: 27).

Oba opisywane programy zostały wyposażone w narzędzia umożliwiające wyrażenie natury zależności pomiędzy kategoriami. Używając funkcji *Relations* w Atlas.ti oraz *Relationships* w NVivo, można określić rodzaj relacji oraz to, czy jest ona jedno czy dwukierunkowa lub po prostu ustalić sam fakt istnienia zależności, bez wskazywania jej kierunku. Dzięki temu uzyskujemy narzędzie wspomagające proces tworzenia hipotez, a więc przechodzenia na wyższy poziom analizy konceptualnej zgodnie z zasadami metodologii teorii ugruntowanej.

Pisanie not teoretycznych

Kolejną kwestią, na jaką należy zwrócić uwagę w kontekście posługiwania się metodologią teorii

ugruntowanej, jest system pisania wszelkiego rodzaju notatek, w tym przede wszystkim not teoretycznych. Noty są zapisem przemyśleń analitycznych związanych z kodami i służą uściśleniu stosowanych kategorii oraz ukierunkowywaniu procesu kodowania. Ponadto, jak podaje Gibbs (2011:68), są również swoistym łącznikiem pomiędzy dwoma etapami kodowania – kodowaniem i pisanem raportu. Natomiast zdaniem Marka Gorzki (2008: 101), pisemne noty, tworząc swego rodzaju myślową i teoretyczną przestrzeń, w której badacz dokonuje konceptualizacji danych, są swego rodzaju narzędziem analizy.

W programach NVivo oraz Atlas.ti rolę not teoretycznych pełnią *memos*, a więc zapisy myśli teoretycznych i koncepcji badacza. *Memos* to zwięzłe notatki badacza zawierające informacje o pomysłach dotyczących całego projektu, konkretnych materiałów bądź też kwestii do dyskusji lub przyszłej interpretacji. Co więcej, badacz może takie notatki modyfikować na każdym etapie procesu analizy. Jednocześnie wszystkie *memos* – z uwagi na ich ważną rolę w procesie analizy – zostały wydzielone w odrębną kategorię elementów projektu (Niedbalski, Ślęzak 2012: 137). *Memos* można także kodować, by uwypuklić istotne dane, które są w nich zawarte lub też w celu skupienia się na nowych wątkach bądź pytaniach, jakie pojawiły się w trakcie prowadzonej analizy. Przy czym w przypadku Atlas.ti, aby móc wykonać powyższe czynności, należy najpierw dokonać konwersji *memo* na zwykły dokument pierwotny. Program NVivo jest w tym zakresie bardziej elastyczny i pozwala na bezpośrednie kodowanie treści zawartych w notach teoretycznych utworzonych w projekcie (Niedbalski, Ślęzak 2012: 137).

Noty dają badaczowi możliwość utrwalania swoich pomysłów analitycznych, ich systematycznego rozwijania, wzbogacania, modyfikowania, a więc jednocześnie prowadzenia pracy analitycznej oraz jej kontroli.

Funkcja *memos* w omawianych programach jest zatem analogiczna do procedury tworzenia not w metodologii teorii ugruntowanej. W ten sposób wyraźnie uwidacznia się także zbieżność funkcji dostępnych w NVivo oraz Atlas.ti z wymogami, jakie przed badaczem stawia wspomniana metodologia badań. Ich właściwe wykorzystanie, a więc sukcesywne i systematyczne tworzenie, może skutecznie przyczynić się do wygenerowania teorii. Nawet bowiem luźne uwagi i ulotne pomysły mogą mieć wpływ na analizę i interpretację, a także na porównanie danych w procesie ich przeglądania (por. Wiltshier 2011; Niedbalski, Ślęzak 2012: 139).

Metoda ciągłego porównywania

Główne zalecenie twórców metodologii teorii ugruntowanej (Strauss, Glaser 1967) polega na prowadzeniu ciągłych porównań, a więc zestawieniu ze sobą fragmentów opisanych podobnym kodem czy porównaniu sposobu zakodowania danego przypadku z innym przypadkiem. W szczególności powinno się poszukiwać pewnych wzorców, dokonywać porównań, formułować wyjaśnienia i budować modele (Gibbs 2011: 141). Metoda ciągłego porównywania polega więc na poszukiwaniu różnic oraz podobieństw pomiędzy fragmentami danych. Jest to zatem konfrontowanie ze sobą różnych fragmentów danych w celu sprawdzenia istniejących między nimi podobieństw bądź wydobycia pew-

nych wyróżniających je cech. Na bazie analizy podobieństw i różnic generowane są coraz ogólniejsze kategorie wydobywające na jaw tkwiące u podłoża zjawisk wymiary jednorodności (*underlying uniformities*) (Gorzko 2008: 86).

Programy NVivo oraz Atlas.ti ułatwiają działania związane z procedurą ciągłego porównywania, a więc kroki niezbędne w procesie budowania teorii (Glaser, Strauss 1967; Konecki 2000). W praktyce, korzystając z opisywanych programów, procedurę porównania wykonuje się, stosując opcje prostego przeglądania zawartości poszczególnych kategorii, generowania zestawień danych w postaci tabel i matryc oraz ich zaawansowanej kwerendy (Niedbalski 2014: 156 i nast.).

W pierwszym przypadku, proces przeglądania zawartości kodów polega na przeglądaniu fragmentów tekstu oraz innych rodzajów danych, które zostały zakodowane danym kodem. W ten sposób uzyskujemy wiedzę na przykład odnośnie tego, jak na dany temat, który został przez badacza ujęty w formie kategorii analitycznej, wypowiadają się poszczególni rozmówcy. Na podstawie przeglądu zawartości kodów możemy dokładnie przyjrzeć się wszystkim fragmentom danych, które zakodowane zostały wybranym kodem. W ten sposób zyskujemy możliwość wglądu w dotychczasowy proces analizy danych pod kątem konkretnych tematów, które reprezentowane są przez określone kategorie. Taka kwerenda, przeprowadzona w sposób skrupulatny, może pomóc w nakreśleniu charakterystyki wybranych wątków w poszczególnych przypadkach, a także stanowić podstawę do sformułowania wyjaśnień interesujących badacza kwestii (Niedbalski 2014: 158–161).

W przeprowadzeniu porównań danych dobrze sprawdzają się również tabele i matryce. Przy czym mają one zupełnie inne zastosowanie i rządzą się inną logiką niż te znane z analizy ilościowej. Ich budowa opiera się zaś na wyszukaniu fragmentów zakodowanego tekstu i ich umieszczeniu, często w streszczonej formie, w poszczególnych polach tabeli (Gibbs 2011: 141). Przyglądając się zawartości poszczególnych pól, można zacząć formułować wyjaśnienia zaobserwowanych różnic bądź podobieństw pomiędzy przypadkami i w ten sposób starać się odnaleźć występujące wzorce.

Spośród dwóch prezentowanych w artykule programów, Atlas.ti i NVivo, zwłaszcza ten ostatni posiada interesujące dla badanych posługujących się MTU narzędzie do generowania matryc (*Framework Matrix*). W praktyce powstają one jako rezultat zestawienia ze sobą między innymi przypadków (w wierszach) z określonymi wątkami i tematami analizowanymi przez badacza (w kolumnach). Dzięki wykorzystaniu tego narzędzia możliwe staje się skondensowanie objętościowo dużych materiałów i ich tematyczne posegregowanie na bardziej przystępne i czytelne fragmenty, które w znacznie prostszy sposób można ze sobą porównywać. Praca z danymi zamieszczonymi w matrycy umożliwia między innymi zapoznanie się z tym, jak przekrojowo w wypowiedziach różnych osób wyglądają te same kwestie, jak o różnych sprawach wypowiada się ta sama osoba czy też, jak kształtują się doświadczenia czy określone zagadnienia względem różnych kategorii rozmówców (Niedbalski 2014: 161).

Również Atlas.ti wyposażony został w narzędzia do tworzenia tabelarycznych zestawień danych. W roli

tej sprawdza się *Codes-Primary Documents Table*. Jest to zbiorcze ujęcie wszystkich kodów występujących w dokumentach pierwotnych projektu. Dzięki niemu możliwe jest sprawdzenie, w jakim stopniu kolejne dokumenty pierwotne (np. wywiady narracyjne) przyczyniają się do pogłębiania wiedzy badacza w zakresie wybranego zagadnienia. Narzędzie to zawiera opcję zliczania częstotliwości występowania dla każdego kodu lub rodziny kodów w dokumentach lub rodzinach dokumentów, a także liczby słów w zakodowanych segmentach oraz dokumentach pierwotnych.

Ponadto, w Atlas.ti mamy do dyspozycji opcję *Co-occurrence Table Explorer*. Na podstawie utworzonego zestawienia możemy określić częstotliwość współwystępowania wybranych kodów. Co ważne, kliknięcie na którąkolwiek z komórek tabeli spowoduje wyświetlenie się listy cytatów zakodowanych przez ujęte w matrycy kody. Liczby, które są widoczne w komórkach, mogą nam pomóc w „rozeznaniu się” w sytuacji współwystępowania zakodowanych tymi samymi kodami fragmentów danych. Nie należy jednak traktować ich w sposób ilościowy, bowiem fakt, że w jednej komórce jest większa cyfra niż w innej, nie musi oznaczać od razu „siły” związku. Jak bowiem wskazywali Strauss i Glaser (1967), częstotliwość współwystępowania nie musi przekładać się na znaczenie kategorii. Aby więc ustalić faktycznie, jak wygląda sytuacja, należy przeprowadzić kolejne kwerendy oraz dokładnie przejrzeć zawartość cytatów w szerszym kontekście ich występowania, a więc powrócić do transkrypcji dokumentów pierwotnych.

Możliwości takie daje również tworzona w programie NVivo matryca kodów (*Matrix Coding Query*),

czyli narzędzie umożliwiające zadanie szeregu pytań na temat wzorców w danych oraz uzyskanie dostępu do ich zawartości. Co warto zaznaczyć, powstała matryca jest wynikiem zastosowania procedury przeszukiwań jednej grupy kodów w odniesieniu do jakiejś drugiej grupy (Gibbs 2011: 241–242). Zarówno w NVivo, jak i w Atlas.ti mamy do dyspozycji całą baterię opcji przeszukiwania pod względem kodów.

W NVivo większość opcji przeszukiwania (zarówno te dotyczące tekstu, jak i kodów) umieszczona jest pod zbiorczą nazwą *Queries* (Zapytania). Aby zatem skorzystać z oferowanych przez program w tym zakresie opcji, należy z menu górnego *Explore* wybrać *New Query*, a z rozwiniętej listy, jedną z wyświetlonych pozycji: *Text Search* (przeszukuje słowa, frazy w tekście), *Word Frequency* (pokazuje częstotliwość występowania słów bądź fraz), *Coding* (odnajduje fragmenty zakodowane kodem bądź kilkoma kodami oraz atrybutami), *Matrix Coding* (tworzy tabelaryczne zestawienie występowania kombinacji kodów, atrybutów oraz innych elementów projektu), *Compound* (pozwala na przeszukiwanie słownikowe z połączeniem przeszukiwania w oparciu o kody), *Coding Comparison* (porównuje kodowanie wykonane przez dwóch użytkowników lub dwie grupy użytkowników), *Group* (odnajduje związane ze sobą elementy projektu, na podstawie kodów, relacji, linków itd.).

Podobnie jak w NVivo, także w programie Atlas.ti do dyspozycji badacza jest rozbudowane narzędzie przeszukiwania danych oraz tworzenia wszelkiego rodzaju ich zestawień. Najważniejsze z nich to *Query Tools*, które w Atlas.ti występuje w formie okna,

w którym dokonuje się wyboru kodów oraz operatorów, jakie służą do ustanowienia różnych sposobów przeszukiwania danych. Generalnie można je pogrupować w trzy główne kategorie. Mamy więc operatory logiczne, semantyczne i przestrzenne/bliskości. Odpowiadają one określonym sposobom przeszukiwania (Niedbalski 2014: 196–197).

Dzięki wykorzystaniu narzędzia zapytań można sprawdzić kodowanie oraz znaleźć nowe wątki, które naprowadzą badacza na kolejne tropy. Można też tworzyć zapytania do poszukiwania odpowiedzi na rodzące się pytania badacza oraz znaleźć pewne wzorce na podstawie powtarzających się słów czy kodów, a także sprawdzić spójność kodowania pomiędzy członkami zespołu. Narzędzia przeszukiwania pod względem kodów mają jeszcze tę istotną zaletę, że mogą wspomóc proces generowania hipotez badawczych, odgrywających niezwykle istotną rolę w badaniach opartych na MTU, bowiem służą budowaniu teorii.

Budowanie i ugruntowywanie hipotez

Oprogramowanie do komputerowego wspomaganie analizy danych jakościowych może być również przydatne do udoskonalania koncepcji teoretycznych, a także tworzenia i „weryfikowania” hipotez. Przy czym nie należy zapominać, że w badaniach jakościowych sprawdzanie hipotez jest procesem, którego nie można sprowadzać do statystycznego testowania. Zamiast „testowania” i „potwierdzania” hipotez (Miles, Huberman 2000: 262) czy ich „weryfikacji” (Strauss, Corbin 1990: 108), proces ich przeglądania oznacza powrót do danych (tj. ponowne czytanie stenogramu bądź notatki terenowej) lub

powrót do terenu badań (tj. prowadzenia nowych obserwacji i wywiadów), aby znaleźć jakieś potwierdzające lub dyskwalifikujące dowody co do słuszności hipotezy. Wynika to między innymi z samej specyfiki hipotez jakościowych, które, zwłaszcza w swej początkowej postaci, są zazwyczaj dość luźnymi „skojarzeniami” badacza. Nie są to konkretne propozycje dotyczące pewnych faktów, ale wstępne i nieprecyzyjne, czasami bardzo niejasne domysły badacza na temat możliwych powiązań pomiędzy kategoriami (Niedbalski 2014: 213).

W obydwu programach dla sprawdzenia intuicji badacza czy inaczej mówiąc, wygenerowanych hipotez, możemy wykorzystać opcje przeszukiwania oparte na podstawie operatorów przestrzennej bliskości bądź operatorów logicznych.

W pierwszym przypadku weryfikacja polega na sprawdzeniu, czy wybrane przez nas kody współwystępują, a więc, czy materiały źródłowe zostały zakodowane w taki sposób, że w pewnej określonej odległości od siebie znajdują się fragmenty danych zakodowane wybranymi kodami. Jeśli tak jest, a sytuacja taka nie jest odosobniona, lecz pojawia się w kolejnych przypadkach, możemy uznać, że (przynajmniej z aktualnej puli danych) hipoteza została potwierdzona. Dzięki temu możliwe jest bowiem odszukanie cytatów przypisanych do kodów nadrzędnych w kierunku podrzędnych (w przypadku zastosowania operatora POD[rzędność]), cytatów przypisanych do kodów od podrzędnych w kierunku nadrzędnych (w przypadku zastosowania operatora NAD[rzędność]) bądź cytatów przypisanych do kodów o tej samej pozycji znaczeniowej (w przypadku zastosowania operatora WSPÓŁ[rzędność]).

Sposobem na weryfikację hipotez, opartym o mechanizm współwystępowania kodów (bądź też ich wzajemnego wykluczania się), jest również skorzystanie z narzędzia przeszukiwania kodów z użyciem operatorów logicznych, takich jak na przykład AND (podaje wyniki przypisane jednocześnie do obu zadanych kodów), OR (podaje wyniki przypisane przynajmniej do jednego z zadanych kodów), AND NOT (podaje wyniki, które nie są przypisane do danego kodu).

Procedura badania hipotezy polega więc na użyciu opcji przeszukiwania kodów, które są używane jako narzędzie heurystyczne. Celem jest bowiem nie weryfikacja poprawności hipotezy, lecz pobranie oryginalnego tekstu zakodowanego przez dane kody i powtórne jego przeanalizowanie.

Tworzenie modeli i diagramów integrujących

Anselm Strauss i Juliet Corbin (1990) sugerują, aby podczas budowania teorii wspierać się na wizualnych reprezentacjach powiązań i zależności istniejących pomiędzy wygenerowanymi kategoriami analitycznymi. Wszelkiego rodzaju wizualizacje w postaci wykresów, diagramów czy sieci są przydatne, aby uporządkować relacje pomiędzy kategoriami, które powstają podczas kodowania osiowego i selektywnego. Przy czym spośród nich najistotniejsze – z perspektywy metodologii teorii ugruntowanej – zdają się być modele (sieci) będące podstawą diagramów integrujących dane. Wszelkiego rodzaju schematy, wykresy czy modele służą wizualnej prezentacji związków i zależności, jakie istnieją pomiędzy składnikami budowanej teorii.

Tworzenie modeli ma kluczowe znaczenie w całym procesie analizy danych, choć szczególnie jest to widoczne w sytuacji krystalizowania się koncepcji badacza, co z kolei ściśle sprzężone jest z wygenerowaniem kategorii centralnej. Ukazanie kategorii oraz powiązań między nimi w sposób zwizualizowany proces ten znacznie ułatwia. A zatem modele odgrywają kluczową rolę podczas kodowania selektywnego, kiedy to ma miejsce krystalizowanie się głównego zjawiska czy procesu, do którego będą się odnosić wszystkie pozostałe, wygenerowane w toku analizy kategorie (por. Niedbalski 2013: 180).

W praktyce każdy model składa się z dwóch zasadniczych elementów, czyli „węzłów” oraz „łączników”. Węzłami sieci (modeli) mogą być najważniejsze elementy projektu. W przypadku NVivo oraz Atlas.ti rodzaj składników zależy od tego, z jakich elementów budowany jest cały projekt. Natomiast łączniki to rodzaje relacji istniejących pomiędzy poszczególnymi węzłami. Węzły w sieci mogą być teoretycznie połączone z dowolną liczbą innych węzłów. Łączniki są zazwyczaj rysowane jako linie między węzłami. Ponadto związki pomiędzy dwoma węzłami mogą być kierunkowe bądź nie. Jeśli jest to połączenie kierunkowe, wówczas jego zwrot zobrazowany jest przez grot strzałki. Przy czym w obrębie powiązań kierunkowych można mówić także o dwóch ich rodzajach: symetrycznych i asymetrycznych (Niedbalski 2014: 218).

Cechą modelu jest możliwość odzwierciedlenia za jego pośrednictwem relacji i powiązań pomiędzy określonymi elementami projektu. W ten sposób

można na bieżąco i w sposób transparentny śledzić, jaki efekt będą miały kolejne etapy analizy. Kiedy jednak badacz będzie chciał utworzyć rodzaj „migawki”, a więc trwale zachować ustawienia danego modelu (sieci), wówczas przyjmie on postać statyczną. Tym samym, jeśli będziemy wprowadzać jakiegokolwiek zmiany i modyfikacje w elementach projektu, owe zmiany nie będą już odzwierciedlane w modelu. Poprzez utworzenie statycznego modelu, badacz może utrwalić swoje pomysły i etapy kreowania nowych idei, do których zawsze może powrócić i sprawdzić, jakie zamiany zostały poczynione oraz jak w danym momencie kształtowały się jego pomysły analityczne (Niedbalski, Ślęzak 2012: 152). Opcja ta może być zatem niezwykle przydatna do dokumentowania kolejnych etapów pracy analitycznej badacza i może stanowić istotną, z perspektywy metodologii teorii ugruntowanej, naturalną historię badania (Bringer, Johnston, Brackenridge 2004).

Pisanie raportu, eksportowanie danych i upowszechnianie wyników badań

Oba programy, NVivo oraz Atlas.ti, wspomagają również pracę badacza na poziomie generowania raportów z przeprowadzonych analiz. Zgodnie ze słowami Gibbsa (2011: 172), dobry, cechujący się refleksyjnością raport z badań powinien mieć wyraźne ugruntowanie w zgromadzonym i przeanalizowanym materiale empirycznym. Głównym sposobem osiągnięcia owego osadzenia w danych jest przedstawienie dowodów potwierdzających wnioski – w postaci cytatów z notatek terenowych, wywiadów lub innych dokumentów zgromadzonych w toku badań.

W opisywanych programach, dzięki takim narzędziom, jak generatory raportów, można: po pierwsze, na bieżąco przeglądać i weryfikować postępy swojej pracy; po drugie, skoncentrować się na wybranych wątkach i tematach, które pojawiają się w projekcie; po trzecie, przedstawić wyniki i postępy pracy analitycznej; po czwarte, pracować z danymi w innych aplikacjach (m.in. Word bądź Excel).

Raporty zawierają informacje podsumowujące projekt, które można nie tylko przeglądać, ale także drukować. Na przykład można sprawdzić postępy kodowania, uruchamiając raport z listą materiałów źródłowych i węzłów, które posłużyły do ich zakodowania. Przy czym warto pamiętać, że raporty nie zawierają zawartości materiałów źródłowych i węzłów (kategorii). Jeśli zatem chcemy zobaczyć taką zawartość, musimy wybrać narzędzia służące do tworzenia kwerendy (Niedbalski 2014: 246 i dalsze).

Warto jeszcze raz zaznaczyć, że zarówno program NVivo, jak i Atlas.ti oferuje możliwość generowania sprawozdań w postaci możliwej do opublikowania w formie strony internetowej. Badacz, który w szybki i łatwy sposób zechce dotrzeć do większej liczby odbiorców – w celu na przykład upowszechnienia wyników swoich badań, może skorzystać z ciekawej opcji tworzenia ministron internetowych (*mini-websites*). Do elementów projektu, które wraz z całą swoją zawartością (a więc powiązaniem, linkami, adnotacjami, *memos* itd.) mogą zostać w ten sposób wyeksportowane i zapisane w formacie .html, należą między innymi wygenerowane kategorie bądź uprzednio zaimportowane

(lub utworzone) w projekcie materiały źródłowe. Wykorzystanie tej funkcji pozwala badaczowi na rozpowszechnienie określonych informacji związanych z projektem oraz umożliwia szerokiemu gronu odbiorców wgląd w aktualny stan zaawansowania prowadzonej analizy danych (Niedbalski, Ślęzak 2012: 154–155).

Podsumowanie

Oprogramowanie NVivo, jak również Atlas.ti wyposażone jest w instrumenty, które znacznie ułatwiają spełnienie wymogów związanych z generowaniem teorii ugruntowanej, dostarczając badaczowi nowych narzędzi dbania o to, by powstająca teoria była dostosowana (*fit*) do danych oraz dająca się modyfikować (*modifiable*) (Glaser 1978; Konecki 2000).

Programy Atlas.ti oraz NVivo pozwalają na wykonywanie określonych działań związanych z procesem analitycznym, które w takim wymiarze oraz w tak relatywnie krótkim czasie nie były możliwe przy zastosowaniu tradycyjnych sposobów prowadzenia badań.

Oba opisywane programy różnią między sobą co do określonych rozwiązań technicznych, które zostały w nich zastosowane, a także „architektury” oprogramowania. Z tego względu sposób, w jaki badacz będzie z nich korzystał, również będzie nieco inny w każdym przypadku. Niemniej jednak to, co wydaje się najważniejsze, to fakt, że przy pomocy jednego, jak i drugiego programu, wykorzystując dostępne w nich narzędzia i ich funkcje, można z powodzeniem realizować badania oparte

na metodologii teorii ugruntowanej i jej procedurach. Przy czym to, w jaki dokładnie sposób owe narzędzia będą wykorzystane, zależy od konkretnego badacza, a niniejszy artykuł pokazuje tylko jedną z możliwych dróg i w żadnym razie wyznacza jedyne wzorca.

Na zakończenie należy jeszcze raz bardzo wyraźnie zaznaczyć, że komputerowe wspomaganie analizy danych jakościowych nie oznacza najlepszego sposobu projektowania i prowadzenia badań, lecz stanowi pewną alternatywę w stosunku do tradycyjnych metod ich wykonywania. To, jakiego wyboru dokona badacz, powinno zależeć od jego osobistych preferencji, rodzaju badań oraz charak-

teru eksplorowanego terenu (Niedbalski, Ślęzak 2012: 161). Najważniejsze wydaje się bowiem nie to, jaki będzie ostatecznie sposób realizacji badań – tradycyjny czy z zastosowaniem CAQDAS – ale kwestia doboru właściwych technik, metod i narzędzi badawczych, które będą dopasowywane do zaplanowanych działań (Seale 2008). Żadne oprogramowanie CAQDA nie wyręczy też badacza i nie wykona analizy w sposób automatyczny (Lonkila 1995). Jeśli zaś rezultaty badań, w których wykorzystano oprogramowanie komputerowe, nie okażą się zadowalające, to będzie to efekt błędów podczas pracy badacza analityka, a nie wad programu (Niedbalski, Ślęzak 2012: 162; por. Bringer i in. 2006; Silverman 2008).

Bibliografia

Bringer Joy D., Johnston Lynne H., Brackenridge Celia H. (2004) *Maximizing Transparency in a Doctoral Thesis I: The Complexities of Writing About the Use of QSR*NVIVO Within a Grounded Theory Study*. „Qualitative Research”, vol. 4, no. 2 s. 247–265 [dostęp 10 października 2012 r.]. Dostępny w Internecie: <<http://qrj.sagepub.com/content/4/2/247>>.

----- (2006) *Using Computer-Assisted Qualitative Data Analysis Software to Develop a Grounded Theory Project*. „Field Methods”, vol. 18, no. 3 s. 245–266 [dostęp 10 października 2012 r.]. Dostępny w Internecie: <<http://fmj.sagepub.com/content/18/3/245>>.

Brosz Maciej (2012) *Komputerowe wspomaganie badań jakościowych. Zastosowanie pakietu NVivo w analizie materiałów nieustrukturyzowanych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1 s. 98–125 [dostęp 20 listopada 2012 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Budziszewska Magdalena (2010) *Możliwości i wykorzystanie programu Atlas.ti w psychologicznych badaniach narracyjnych* [w:] Maria Straś-Romanowska, Bogna Bartsch, Magdalena Żurko, red., *Badania narracyjne w psychologii*. Warszawa: Wydawnictwo Psychologii i Kultury, s. 83–100.

Fielding Nigel (2007) *Computer Applications in Qualitative Research* [w:] Paul Atkinson et al., eds., *Handbook of Ethnography*. Los Angeles, London, New Delhi, Singapore: Sage, s. 453–467.

Gibbs Graham (2011) *Analizowanie danych jakościowych*. Przełożyła Maja Brzozowska-Brywczyńska. Warszawa: Wydawnictwo Naukowe PWN. Glaser Barney G. (1978) *Theoretical Sensitivity*. San Francisco: The Sociology Press.

Glaser Barney G., Holton Judith (2004) *Remodeling Grounded Theory*. „Forum Qualitative Sozialforschung/ Forum: Qualita-

tive Social Research”, vol. 5, no. 2, Art. 4 [dostęp 20 listopada 2012 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/607/1315>>.

Glaser Barney G., Holton Judith (2010) *Przemodelowywanie teorii ugruntowanej*. Przełożyli Marek Gorzko i Łukasz Pyfel. „Przegląd Socjologii Jakościowej”, t. 6, nr 2, s. 81–102 [dostęp 15 września 2013 r.]. Dostępny w Internecie: <http://www.qualitativesociologyreview.org/PL/Volume13/PSJ_6_2_Glaser_Holton.pdf>.

Glaser Barney G., Strauss Anselm L. (1967) *The discovery of grounded theory. Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Gorzko Marek (2008) *Procedury i emergencja. O metodologii klasycznych odmian teorii ugruntowanej*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Jones Michael, Diment Kieren (2010) *The CAQDA Paradox: A divergence between research method and analytical tool*. „The International workshop on Computer-Aided Qualitative Research Asia: (CAQRA2010)”. The Netherlands: Merlien Institute.

Kelle Udo (2005) *Komputer-Assisted Qualitative Data Analysis* [w:] Clive Seale i in., eds., *Qualitative Research Practise*. London, Thousand Oaks, New Delhi: Sage, s. 473–489.

Konecki Krzysztof (2000) *Studia z metodologii badań jakościowych. Teoria ugruntowana*. Warszawa: Wydawnictwo Naukowe PWN.

Lewins Ann, Christian Silver (2004) *Choosing CAQDAS software. CAQDAS Networking Project*. Guildford: University of Surrey.

Lonkila Markku (1995) *Grounded theory as an emerging paradigm for computer-assisted qualitative data analysis* [w:] Udo Kelle, ed., *Computer-Aided Qualitative Data Analysis*. London: Sage, s. 41–51.

Miles Matthew B., Huberman Michael A. (2000) *Analiza danych jakościowych*. Przełożył Stanisław Zabielski. Białystok: Transhumana.

Niedbalski Jakub (2013) *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub (2014) *Komputerowe wspomaganie analizy danych jakościowych. Zastosowanie oprogramowania NVivo i Atlas.ti w projektach badawczych opartych na metodologii teorii ugruntowanej*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1 s. 126–165 [dostęp 20 listopada 2012 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Saillard Elif K. (2011) *Systematic Versus Interpretive Analysis with Two CAQDAS Packages: NVivo and MAXQDA*. „Forum: Qualitative Social Research”, vol. 12, no. 1 [dostęp 30 września 2012 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1518>>.

Salinger Stephan, Plonka Laura, Prechelt Lutz (2007) *A Coding Scheme Development Methodology Using Grounded Theory for Qualitative Analysis of Pair Programming*, „Human Technology”, vol. 4, no. 1, s. 9–25.

Schönfelder Walter (2011) *CAQDAS and Qualitative Syllogism Logic—NVivo 8 and MAXQDA 10 Compare*. „Forum: Qualitative Social Research”, vol. 12, no. 1 [dostęp 30 września 2011 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/article/view/1514>>.

Seale Clive (2008) *Wykorzystanie komputera w analizie danych jakościowych* [w:] David Silverman, red., *Prowadzenie badań jakościowych*. Warszawa: Wydawnictwo Naukowe PWN, s. 233–256.

Silverman David (2008) *Interpretacja danych jakościowych*. Przełożyła Joanna Ostrowska. Warszawa: Wydawnictwo Naukowe PWN. Strauss Anselm L. (1987), *Qualitative Analysis for Social Scientists*. Cambridge: University Press Cambridge.

Strauss Anselm L., Corbin Juliet (1990) *Basics of Qualitative Research*. London–New Delhi: Sage.

Wiltshier Fiona (2011) *Researching With NVivo*. „Forum: Qualitative Social Research”, vol. 12, no. 1 [dostęp 30 września 2012 r.]. Dostępny w Internecie: <<http://www.qualitative-research.net/index.php/fqs/issue/view/36>>.

Cytowanie

Niedbalski Jakub (2014) *Zastosowanie oprogramowania Atlas.ti i NVivo w realizacji badań opartych na metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 60–80 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

The Application of Atlas.ti and NVivo Software in Conducting Researches Based on Grounded Theory Methodology

Abstract: This article raises a topic of special software applied to support the analysis of qualitative data in research which is based on the procedures of grounded theory methodology (GTM). The purpose of this article is to demonstrate what kind of relations occur between the methodology procedures of grounded theory and two popular programs of CAQDA group: NVivo and Atlas.ti. This article is intended to show in what ways these two programs can be used to provide a GT-based analysis. In the article, there is a demonstration of both technical and applicable possibilities of NVivo and Atlas.ti software. Moreover, this article points out a degree of adequacy of technical solutions applied in both programs in order to meet the requirements of grounded theory methodology, as well as some restrictions and barriers which can be encountered by a researcher who uses a particular computer program in GT-based research.

Keywords: Computer Assisted Qualitative Data Analysis Software, New Technologies, Social Research Methods, Methodology of Grounded Theory, NVivo, Atlas.ti

Jacek Burski
Uniwersytet Łódzki

Kamil Brzeziński
Uniwersytet Łódzki

DOI: <https://doi.org/10.18778/1733-8069.10.2.05>

Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

Abstrakt Celem niniejszego artykułu jest przedstawienie podstawowych założeń metodologicznych i teoretycznych projektu „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” oraz zaprezentowanie możliwości wsparcia realizacji procesu badawczego poprzez wykorzystanie jednego z programów CAQDAS.

W artykule pokazana zostanie próba przedstawienia wstępnych wniosków z wykorzystania programu komputerowego QDA Miner, którego wydawcą jest firma Provalis, w analizie historycznych wypowiedzi prasowych. Zaprezentowane zostaną możliwości zastosowanego oprogramowania, jak i jego wpływ na przebieg całego procesu badawczego. Przedstawiony zostanie również sposób, w jaki wybór i wykorzystanie danego programu determinują decyzje dotyczące analizowanego materiału empirycznego, jego selekcję, kategoryzację i deskrypcję. Dodatkowym celem artykułu jest poszerzenie wiedzy metodologicznej i praktycznej z zakresu wykorzystywania wsparcia oprogramowania komputerowego w jakościowej analizie danych.

Artykuł został podzielony na kilka części, w pierwszej z nich przedstawiono krótki zarys historii wykorzystywania oprogramowania komputerowego do analiz jakościowych. Część druga poświęcona została podstawowym założeniom projektu badawczego „Cztery dyskursy o nowoczesności...”. Następnie omówiono sposób doboru oprogramowania odpowiadającego wymogom projektu, a w dalszej kolejności zaprezentowano pierwsze próby jego wykorzystania w ramach badania oraz jego wady i zalety.

Słowa kluczowe analiza jakościowa wspomagana komputerowo, QDA Miner, Łódź, dyskurs, nowoczesność

Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

Początki CAQDAS

Przez długi okres specyficzne cechy badań jakościowych, takie jak między innymi chęć dokładnego przyjrzenia się światu zewnętrznemu oraz wyjaśnienia zjawisk społecznych z perspektywy wewnętrznej (Flick 2010), sprawiały, iż wydawało się, że wykorzystanie programów komputerowych w pracy badacza jakościowego było zbędne, a nawet niemożliwe. Wykorzystanie komputera w pro-

Jacek Burski, mgr, doktorant w Katedrze Socjologii Kultury Instytutu Socjologii Uniwersytetu Łódzkiego. Jego zainteresowania koncentrują się wokół problematyki socjologii sportu, socjologii biografii oraz metodologii badań jakościowych – w tym problemów z wykorzystaniem oprogramowania komputerowego. Dodatkowo w kręgu jego działań badawczych leżą kwestie związane z problematyką nowoczesności.

Adres kontaktowy:

Katedra Socjologii Kultury
Instytut Socjologii Uniwersytetu Łódzkiego
ul. Rewolucji 1905 r. 41, 90-214 Łódź
e-mail: jacekburski84@gmail.com

Kamil Brzeziński, mgr, asystent w Katedrze Socjologii Wsi i Miasta Uniwersytetu Łódzkiego. Główne obszary jego zainteresowań naukowych to socjologia miasta, metody oraz metodologia badań społecznych.

Adres kontaktowy:

Katedra Socjologii Wsi i Miasta
Instytut Socjologii Uniwersytetu Łódzkiego
ul. Rewolucji 1905 r. 41, 90-214 Łódź
e-mail: kamilbrzezinski84@gmail.com

cesie badawczym kojarzone było zwykle z opracowywaniem zestawień i analiz statystycznych. Niemniej jednak szybki rozwój technologiczny, a w tym przede wszystkim pojawienie się programów typu CAQDAS (Computer-Assisted Qualitative Data Analysis Software), sprawił, że w chwili obecnej badacze jakościowi zyskali potężne narzędzie wspierające przeprowadzenie analizy danych jakościowych przy pomocy oprogramowania komputerowego.

Jak zauważa w jednym ze swoich artykułów Jakub Niedbalski (2012), metody wykorzystywane przez socjologów jakościowych przechodzą nieustanną ewolucję w zakresie realizacji i organizacji badań. Jedną z głównych przyczyn stałych zmian w realizacji i organizacji badań jakościowych upatrywać należy w rozwoju procesów komputeryzacji i informatyzacji oraz stopniowym upowszechnianiu się nowoczesnych technologii w naukach społecznych (Niedbalski, Ślęzak 2012). Jak dodaje Niedbalski,

mija czas kartki i ołówka, a nowoczesne technologie coraz silniej oddziałują na proces badawczy poprzez dostarczenie nowoczesnych narzędzi metodologicznych, takich jak specjalistyczne programy komputerowe. (2012: 220)

Warto jednak zaznaczyć, że wykorzystywanie nowoczesnych technologii w projektach badawczych nie rozpoczęło się jednocześnie we wszystkich dziedzinach naukowych. W pierwszej kolejności potencjał związany z komputeryzacją doceniony został przez przedstawicieli nauk ścisłych i przyrodniczych (Niedbalski, Ślęzak 2012). Zapewne między innymi ze względu na przedmiot

zainteresowań badawczych, nowoczesne technologie w naukach społecznych pojawiły się zdecydowanie później. Co więcej, początkowo komputery wykorzystywane były jedynie przez przedstawicieli metod ilościowych, którzy używali specjalnych programów do analiz statystycznych. Przykładem jednego z takich programów może być powszechnie znany i wykorzystywany przez socjologów oraz psychologów SPSS (Statistical Package for the Social Sciences) – pakiet statystyczny, który na rynku dostępny jest już niemal od 40 lat. Pakiet ten stosowany był już w okresie tak zwanych komputerów typu *mainframe*, co wymagało od użytkownika znajomości języka poleceń. Wraz z rozwojem technologii oraz komputerów, w tym między innymi pojawieniem się tak zwanych okienkowych systemów operacyjnych, modyfikacjom ulegał także SPSS. Zastosowanie intuicyjnego, graficznego interfejsu w znacznym stopniu ułatwiło korzystanie z programu, aczkolwiek część bardziej zaawansowanych badaczy i analityków nadal preferuje korzystanie z języka poleceń. Oprócz badaczy ilościowych już od początku lat sześćdziesiątych XX wieku potencjał nowych technologii w postaci komputerów *mainframe* dostrzegali również badacze literatury, którzy używali komputerów do liczenia częstotliwości poszczególnych słów w analizowanych tekstach (Seale 2009). W tym początkowym okresie wykorzystywania wsparcia komputerowego w pracach badawczych jednym z podstawowych czynników ograniczających ich powszechne zastosowanie były koszty związane z zakupem komputerów osobistych oraz czasochłonność analiz. Kolejną barierą była kwestia związana z pewnością i wiarygodnością wyników analiz prowadzo-

nych z wykorzystaniem komputera. Jak zauważa Seale (2009), procedury statystyczne były dobrze opisane, co ułatwiało ich wprowadzenie do programów komputerowych, a rezultaty analiz nie budziły większych wątpliwości. Zdecydowanie subtelniejsze procedury postępowania badacza jakościowego nie nadawały się do tak łatwej implementacji do oprogramowania komputerowego. Niemniej jednak rozwój technologii ostatecznie wpłynął również na prowadzenie tego rodzaju badań. Jednym z przykładów może być wykorzystywanie sprzętu nagrywającego, co wpłynęło zarówno na sposób gromadzenia materiałów empirycznych, jak również na sposoby ich analizowania. Gibbs (2011) zauważa, że łatwość uzyskania pełnego zapisu wywiadu bądź rozmowy umożliwiła zdecydowanie dokładniejsze badanie tego, co i jak zostało powiedziane. Ponadto bez możliwości nagrywania dźwięku, a następnie przygotowywania pełnych transkrypcji prowadzenie analizy dyskursu czy analizy konwersacyjnej byłoby niesłychanie trudne, o ile w ogóle niemożliwie. Pojawienie się i upowszechnienie profesjonalnych urządzeń nagrywających wpłynęło znacząco na sposoby zbierania materiałów, jednakże nie wspierało procesów analitycznych, tak jak to miało miejsce w przypadku pakietów statystycznych wykorzystywanych przez badaczy ilościowych. Na podobną zmianę w przypadku badań jakościowych trzeba było poczekać do lat osiemdziesiątych XX wieku, w których pojawiły się pierwsze programy komputerowe wspomagające analizy danych jakościowych – tak zwane CAQDAS (Gibbs 2011). To swego rodzaju opóźnienie w wykorzystywaniu komputerów przez badaczy jakościowych, oprócz wyżej wspomnianych

barier w postaci cen komputerów oraz czasochłonności analiz, wiązało się również z nieukrywana niechęcią tych badaczy w stosunku do nowoczesnych technologii, które utożsamiano przede wszystkim z badaniami ilościowymi. Prace i badania bazujące na statystyce postrzegano jako powierzchowne i zdehumanizowane, gdyż w większym stopniu koncentrowały się one na kontroli, wymaganiach technicznych niż na istotnych społecznie problemach (Seale 2009). Komputer wykorzystywany przez badaczy ilościowych symbolizował wszystkie te negatywne cechy, chcąc zatem niejako przewyciężyć dominującą ortodoksję narzuconą przez zwolenników metod ilościowych, niewskazane było używanie stosowanych przez „ilościowców” rozwiązań. Z czasem, gdy zaczęły pojawiać się kolejne pokolenia badaczy jakościowych oraz powstawały coraz efektywniejsze narzędzia wspierające procesy analizy jakościowej, stosowanie specjalistycznego oprogramowania komputerowego upowszechniło się (Niedbalski, Ślęzak 2012). W tym miejscu warto wspomnieć, jak zmieniały się i rozwijały programy komputerowe do jakościowej analizy danych. Początkowo programy te umożliwiały jedynie ilościową analizę treści: zliczanie liczby poszczególnych słów występujących w tekście; w latach osiemdziesiątych XX wieku możliwe było nadawanie kodów poszczególnym fragmentom tekstu, a w latach dziewięćdziesiątych XX wieku programy te dawały już możliwość tworzenia powiązań między poszczególnymi kodami, a nawet konstruowania teorii (Gibbs 2011; Niedbalski, Ślęzak 2012). Aktualnie istnieje wiele firm, które każdego roku wypuszczają nowe, ulepszone wersje swoich produktów, dające badaczowi coraz więcej możliwości analitycznych.

Do podstawowych właściwości większości dostępnych programów można zaliczyć:

- importowanie i wyświetlanie różnych formatów plików tekstowych;
- konstruowanie listy kodów, w tym także w postaci hierarchicznej;
- wyszukiwanie poszczególnych, zakodowanych partii tekstu;
- analizowanie zakodowanych części tekstu w kontekście oryginalnych dokumentów;
- pisanie not, komentarzy, które można wiązać z wybranymi kodami lub dokumentami (Gibbs 2011).

Programy CAQDAS sprawiają ponadto, że prowadzenie analiz staje się łatwiejsze, szybsze oraz dokładniejsze, badacz nie musi odręcznie zapisywać poszczególnych kodów, wykorzystywać kolorowych markerów czy wycinać poszczególnych fragmentów tekstu z transkrypcji, jak to miało miejsce w przeszłości. Oprogramowanie CAQDAS eliminuje również konieczność kopiowania ogromnej liczby stron między poszczególnymi plikami. Badacz w jednym miejscu może z łatwością podzielić zebrany materiał na wybrane kategorie, zakodować fragmenty poszczególnych tekstów i tak dalej. Z kolei możliwości stworzenia hierarchicznej listy kodów oraz wyszukiwania poszczególnych fraz i wyrazów sprawiają, że badacz posiada większą kontrolę nad całym procesem analitycznym, minimalizując przy tym możliwość pominięcia pewnego fragmentu, co tym samym zwiększa wiarygodność

całego badania. Co więcej, na klarowność procesu badawczego znaczny wpływ mają również możliwości kompleksowego zarządzania znaczną liczbą dokumentów, kodów, not i komentarzy. Jak podkreśla Gibbs (2011), wstępnym warunkiem prawdziwie efektywnej analizy jakościowej jest wydajne i systematyczne zarządzanie danymi. Podsumowując, można stwierdzić, że programy z rodziny CAQDAS zwiększają również dyscyplinę analizy oraz kontrolę badacza nad całym procesem, a tym samym ułatwiają wykazanie, że efekty pracy badaczy oparte są na rygorystycznych procedurach, co w znacznym stopniu przyczynić się może do zwiększenia zaufania w stosunku do ich tekstów (Seale 2009). Kolejną zaletą jest zwiększenie rzetelności pracy zespołowej. Programy CAQDAS pomagają członkom zespołu w ustaleniu, czy dokonali interpretacji fragmentów w ten sam sposób (Durkin 1997 za Seale 2009).

Oprócz powyżej wymienionych zalet i atutów programów wspierających jakościową analizę danych w literaturze przedmiotu można się spotkać również z pewnymi zarzutami i niebezpieczeństwami związanymi z wykorzystywaniem tego typu programów. Poza wyżej wspomnianym – negatywnym – postrzeganiem i łączeniem komputerów wyłącznie z metodami ilościowymi wśród wad wymienia się między innymi zapośredniczony przez program dostęp do materiału badawczego. Jak dodaje Gibbs (2011), badacze, którzy prowadzili analizy na papierze, mieli poczucie bliższej styczności z wypowiedziami respondentów niż ci, którzy korzystali z komputerów. W tym miejscu należy jednak rozważyć, czy dla badacza jakościowego większego znaczenia nie ma możliwość samodzielnego prze-

prowadzenia fazy terenowej – zebrania materiału, przeprowadzenia wywiadów niż możliwość prowadzenia analizy na papierze. Coraz częściej spotkać się można, zwłaszcza w dużych projektach badawczych, że kto inny przeprowadza wywiady, inna osoba z kolei odpowiada za opracowanie transkrypcji, a jeszcze inna za analizę zebranego materiału. Sytuacja taka może pozbawić badacza poznania pełnego kontekstu sytuacji, a także w zdecydowanie większym stopniu styczności z osobami badanymi. W tym kontekście dostęp do materiału zapośredniczony przez monitor komputera nie wydaje się być kluczowym problemem dla całego procesu badawczego.

Jako jedną z wad wymienia się także zbyt silny wpływ metodologii teorii ugruntowanej. W tym przypadku jako przykład dowodzący słuszności tego zarzutu mogą posłużyć programy Ethnograph i Atlas.ti, które zostały zaprojektowane z myślą o tej konkretnej metodologii (Seale 2009; Gibbs 2011). Jednakże należy zaznaczyć, że na rynku dostępna jest coraz większa liczba różnych programów do analizy jakościowej, w tym tych, na które metodologia teorii ugruntowanej nie wywarła tak silnego wpływu jak w przypadku programu Atlas.ti. Oprogramowaniu CAQDAS zarzuca się dodatkowo zbyt silną koncentrację na procedurach kodowania i wyszukiwania danych (Gibbs 2011). Warto jednak zauważyć, że większość wydawców programów do jakościowej analizy danych każdego roku wydaje nowe wersje własnych produktów, wyposażone w coraz nowsze i bardziej zaawansowane narzędzia, które nie ograniczają się jedynie do kodowania. W tym miejscu nie do końca uzasadniony wydaje się kolejny zarzut mówiący, że operacje wy-

konywane w ramach jakościowej analizy danych mogą zostać przeprowadzone w odpowiednim edytorze tekstu bez konieczności nabywania profesjonalnego i drogiego oprogramowania CAQDAS. Niemniej jednak zgodzić się można, że w przypadku niewielkiej ilości materiału w postaci kilku wywiadów, zakup profesjonalnego programu do analizy może nie być w pełni uzasadniony, a tym bardziej opłacalny. Oprócz edytora tekstu, wykorzystać można jeden z programów bezpłatnych, na przykład IHMC Cmap Tools, TranscriberAG, Compedium, Mind42 i tym podobne. Oczywiście należy zaznaczyć, że programy te mogą wspomóc proces badawczy, tworzenie map pojęciowych, aczkolwiek jako oprogramowanie free software'owe mają również znaczne ograniczenia. W kontekście wad i ograniczeń programów CAQDAS podkreśla się również, że oprogramowanie tego typu może okazać się zupełnie bezużyteczne dla badaczy prowadzących analizę konwersacyjną krótkiego wycinka rozmowy czy też analizę dyskursu niewielkiego fragmentu tekstu (Seale 2009). Niezależnie od powyższych, wad środowisko badaczy jakościowych stopniowo przestawało postrzegać komputer jako narzędzie przydatne tylko „ilościowcom”, a zainteresowanie możliwościami wsparcia prac badawczych poprzez wykorzystanie nowoczesnych technologii wzrastało. Coraz więcej osób nie postrzega pracy przy pomocy programów CAQDAS jako znacznego zapośredniczenia w dostępie do materiału, a raczej docenia szybkość i łatwość analiz. Ponadto coraz szersza i bardziej zróżnicowana oferta tego typu oprogramowania sprawia, że w chwili obecnej trudno również utrzymać argument mówiący, że programy CAQDAS przydatne są głównie w sytuacji, gdy korzystamy z metodologii teorii

ugruntowanej. W przypadku projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” specyfika projektu, która została przedstawiona w dalszej części artykułu, wymagała wykorzystania profesjonalnego oprogramowania komputerowego umożliwiającego przeprowadzenia zarówno analizy jakościowej, jak i ilościowej.

Założenia metodologiczno-teoretyczne projektu „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

Zanim przejdziemy do właściwej części tekstu, w której postaramy się wnikliwie przeanalizować relację pomiędzy komputerowym wsparciem w postaci aplikacji służącej do wspomaganiania analizy danych a samym procesem badawczym i zebraniem materiałem empirycznym, chcemy przedstawić czytelnikowi bardziej dokładnie temat i założenia projektu badawczego, który jest podstawą artykułu.

Badanie zostało formalnie rozpoczęte w 2012 roku, choć prace nad założeniami trwały już wcześniej. Zespół badawczy zdobył finansowanie w ramach konkursu OPUS, którego fundatorem jest Narodowe Centrum Nauki. Tytuł projektu: „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” bezpośrednio wskazuje na problematykę, która interesowała badaczy.

Przede wszystkim w centrum uwagi leżał temat nowoczesności oraz jej różnych oblicz, przejawiających

się w określonym miejscu na społeczno-ekonomicznej mapie świata – Europie Środkowo-Wschodniej. Przypadkiem wybranym przez autorów projektu jest Łódź – miasto – rówieśnik rewolucji przemysłowej (a właściwie tej jej wersji, która dotknęła polskie społeczeństwo w XIX i XX wieku). Wybór takiego przedmiotu badania ma w założeniu skutkować powstaniem wielopoziomowego opisu studium przypadku, charakterystycznego między innymi ze względu na swoje umiejscowienie w ramach klasycznego podziału na centrum i peryferie, w którym Łódź na ziemiach polskich była jednym z niewielu przypadków rozwoju podobnego do przemian na zachodzie Europy. Z drugiej strony, ze względu na swoje położenie geopolityczne i etniczny tygiel, żywy w jej początkach i szczytach rozwoju, trudno uznać ją za typ idealny nowoczesnego polskiego miasta. Pytaniem, które zostawimy na boku analiz, jest, czy w ogóle taki typ idealny można wskazać bądź opisać. Specyficzna historia społeczna i ekonomiczna przedmiotu badania ma stanowić podłoże, na którym badacze mają prowadzić analizy nakierowane na problem mechanizmów leżących u podstaw tworzenia się dyskursów nowoczesności, w których pojęcie miasta stanowi klucz do ich zrozumienia. Głównym celem projektu jest przyjrzenie się, w jaki sposób miasto było przedmiotem i podmiotem modernizacji intensyfikującej cechy miasta nowoczesnego. Mechanizmy konstruowania wskazanych powyżej „dyskursów nowoczesności” są analizowane na podstawie materiału pochodzącego z czterech przełomowych momentów rozwoju Polski i Łodzi:

- odzyskanie niepodległości po I wojnie światowej,
- nowy porządek polityczny po II wojnie światowej,
- upadek realnego socjalizmu w 1989 roku.

Przechodząc do bardziej szczegółowego opisu założeń projektu, trzeba podkreślić, że w początkowej fazie miał on na celu operacjonalizację teorii-kulturowych kategorii opisu, by wykorzystać je w socjologicznych badaniach dyskursu poprzez połączenie analiz ilościowych (leksykometria) i jakościowych (badanie konstruktów narracyjnych i semantycznych). Wraz z rozwojem prowadzonych prac oraz problemami, które badacze napotkali przy próbach spełnienia założonych celów, część tych założeń uległa okrojaniu. Element leksykometryczny z powodów, które zostaną bardziej szczegółowo opisane w dalszej części tekstu, został zmarginalizowany. W orbicie zainteresowań nadal pozostaje jednak rozpoznanie sposobów definiowania nowoczesności jako ogólnego pojęcia organizującego dyskurs na temat miasta. Przyjęta hipoteza zakłada, że przedmiot dyskursu, jakim jest miasto (Łódź) i jego reprezentacje, pozostaje w związku z rozmaicie definiowaną nowoczesnością. Chcąc zakreślić bardziej zamknięty obszar badania, przyjęto, że wskazane powyżej momenty przełomowe dają szansę na dostęp do bogatszego zasobu interesujących nas reprezentacji. Uszczegóławiając, analizie poddane zostaną następujące rodzaje materiałów: prasa (łódzka i ogólnopolska); oficjalne wystąpienia przedstawicieli władz samorządowych, przedstawicieli lokalnej elity;

dokumenty wytworzone przez władze lokalne; druki ulotne; publicystyka; broszury; źródła normatywne; dokumentacja wytworzona przez partie polityczne i organizacje społeczne – zgodnie z ich występowaniem w niewielkiej odległości czasowej od podanych przełomów historycznych.

Od strony teoretycznej poszukiwano inspiracji zarówno wśród koncepcji odnoszących się do problemu nowoczesności (wśród których należałoby wymienić przede wszystkim propozycje Marshalla Bermanna czy Agnes Heller), jak również – przy wyborze ramy metodologicznej – posłkowano się dokonaniem analityków dyskursu. W związku z tym, że badanie koncentruje się przede wszystkim na przyczynach wytwarzania, kontestacji i transformacji dyskursów, postanowiliśmy wykorzystać koncepcję zaproponowaną przez tak zwaną post-strukturalistyczną teorię dyskursu (Howarth 2008), która korzystając z ontologicznych założeń na temat dyskursywnej rzeczywistości, pozwala na analityczne spojenie koncepcji nowoczesności jako niedomkniętego projektu wraz z politycznym charakterem debaty na temat miasta. W tym sensie badaczy interesują momenty, kiedy debata ta, wraz ze zmianą warunków społeczno-historycznych, powinna zostać przeddefiniowana na nowo.

W fazie projektowania badania założyliśmy, że ze względu na rozmiar antycypowanego korpusu danych koniecznością (i zarazem szansą) będzie wsparcie klasycznej analizy danych technikami komputerowymi, które pozwoliłyby na bardziej efektywne opisywanie, kategoryzowanie i analizowanie zebranego materiału. Silnym bodźcem do postawienia takiego założenia na pierwszych etapach

konceptualizacji problemu i metody była chęć połączenia jakościowej optyki z wyraźnym akcentem ilościowym. Nacisk miał być położony na wprowadzenie analiz leksykometrycznych, które byłyby niemożliwe bez wsparcia komputerowego. Wraz z rozwojem projektu i podejmowanymi decyzjami natury metodologicznej i logistycznej – będąc jednocześnie w środkowej fazie zbierania i opisywania materiału – możemy powiedzieć, że znacznie zmalała rola leksykometrii w procesie badawczym. Na przeszkodzie stanęły przede wszystkim ograniczenia związane z charakterem materiału badawczego. Teksty pochodzące z różnych epok historycznych nie tylko różnią się na poziomie gramatycznym, ale również znaczeniowym. Biorąc dla przykładu pojęcie z zakresu polityki, która ze względu na swój wpływ na historię miasta stanowi istotne pole analizy, musimy pamiętać o historycznej ewolucji idei i konceptów. Przy jednolitości materiału pod względem językowym można by założyć, że tego rodzaju problem powinien raczej stanowić ciekawe wyzwanie badawcze. Jednakowoż, ze względu na ograniczenia językowe programu oraz wagę innych zadań badawczych, na ten moment analizy leksykometryczne mają mieć¹ raczej charakter wspomagający dla głównych działań badawczych o charakterze jakościowym.

Uzasadnienie wyboru QDA Minera

Tak jak zaznaczono w poprzedniej części artykułu, jednym z pierwszych, a zarazem kluczowym etapem

¹ Należy zaznaczyć, że ze względu na fakt, iż projekt jest cały czas w trakcie realizacji, część zadań nie została jeszcze wykonana. Obecnie zespół badawczy jest w trakcie fazy zbierania materiału, a część analityczna zostanie przeprowadzona w późniejszym terminie.

umożliwiającym prawidłową realizację opisanego powyżej projektu był wybór odpowiedniego oprogramowania komputerowego, umożliwiającego uzyskanie zakładanych we wniosku o dofinansowanie rezultatów badawczych. Podstawowym czynnikiem wpływającym na wybór danego programu była specyfika projektu: jego cele, problematyka i przedmiot badań. Głównym celem projektu jest zbadanie, w jaki sposób w czterech różnych okresach historycznych kształtowała się wyobraźniowa wizja Łodzi jako miasta nowoczesnego, znajdująca odzwierciedlenie w dyskursie, rekonstruowana zgodnie z wyobrażeniami na temat obowiązujących w danym momencie historycznym kryteriów nowoczesności. Powyższy cel ma zostać osiągnięty na podstawie zebranego materiału empirycznego, głównie w postaci prasy, dla każdego z czterech analizowanych okresów. Po zdigitalizowaniu, zebrane materiały zostaną w dalszej kolejności poddane pogłębionej analizie jakościowej, poprzedzonej analizą ilościową. Początkowo zakładano, że analiza ilościowa będzie miała charakter badania leksykometrycznego: analiza częstotliwości, analizy charakterystyk danych subkorpusów i analizy kolokacji oraz współwystępowania słów, co umożliwi badanie większej ilości danych, a także rozpoznanie momentów zmian dyskursu w czasie. W tym przypadku ważne z punktu widzenia założeń projektowych było, aby nabyty program nie korzystał z predefiniowanych kategorii wyrazów, angielskich thesaurusów, stemmingu i lematyzacji.

Analiza jakościowa z kolei wykorzystywać będzie elementy analizy struktur narracyjnych i inspiracje krytycznej analizy dyskursów (badanie diachronicznej zmiany dyskursu). Drugim czynnikiem determinującym wybór oprogramowania był interdyscyplinarny charakter zespołu badawczego, w skład

którego oprócz socjologów, zajmujących się dotychczas różnymi polami badawczymi, wchodzi także historycy oraz kulturoznawcy. Ta swego rodzaju różnorodność zespołu ma pozwolić na korzystanie z najlepszych stron wymienionych dyscyplin, a także przewyższenie ich ograniczenia. W kontekście wyboru stosownego oprogramowania istotne zatem było dobranie programu, który umożliwiałby efektywną pracę zespołu oraz pozwalał na scalenie efektów pracy poszczególnych badaczy, a także dawał członkom zespołu możliwość ustalenia, czy dokonywane przez nich interpretacje poszczególnych fragmentów są spójne. Kolejnym czynnikiem były wynikające z budżetu projektu ograniczenia finansowe oraz chęć uzyskania jak największej liczby licencji, co zapewniałoby możliwość jednoczesnej pracy kilku osobom z zespołu.

Pierwszym napotkanym problemem w wyborze oprogramowania umożliwiającego badanie zgodnie z zapisami z wniosku o dofinansowanie była niska znajomość programów z rodziny CAQDAS przez członków zespołu badawczego. W związku z powyższym zespół został podzielony na podgrupy, które zostały oddelegowane do zebrania informacji na temat dostępnych na rynku programów i zweryfikowania ich pod kątem przydatności dla planowanego badania. Każdy z podzespołów korzystał z bezpłatnych wersji trialowych oraz tutorialowych. Efektem pracy poszczególnych podgrup było przygotowanie sprawozdań, które wskazywały mocne i słabe strony poszczególnych programów. W sumie weryfikacji poddano następujące programy: Atlas.ti, MAXqda, Nvivo, QDA Miner, Leximancer oraz Socjolog 1.5. Ponadto zespół projektowy skonsulto-

wał się w sprawie doboru programu z osobami, które w przeszłości korzystały już z oprogramowania CAQDAS i mogły wskazać, który z analizowanych produktu byłby najlepszym rozwiązaniem, biorąc pod uwagę założenia projektowe. Po przeanalizowaniu sprawozdań oraz sugestii od praktyków, mając na uwadze trzy wcześniej przedstawione czynniki w postaci planowanej metodologii, specyfiki i liczebności zespołu oraz zabudżetowanych środków, zdecydowano się na zakup programu QDA Miner wraz z dodatkiem WordStat. W przypadku programów Atlas.ti, MAXqda oraz Nvivo uznano, że programy te w pełni umożliwiałyby przeprowadzenie jakościowej analizy zebranego materiału, aczkolwiek niezbędne byłoby zakupienie programu dedykowanego analizie ilościowej. Z kolei w odniesieniu do programu Leximancer, który mógłby zostać wykorzystany do ilościowej analizy tekstu, pojawiły się obawy, że może się on okazać zbyt zautomatyzowany i może nie poradzić sobie ze specyfiką języka polskiego. Co więcej, specyfikacje przedstawionych programów sprawiały, że niezbędne byłoby nabycie zarówno programu dedykowanego *stricte* analizie jakościowej, jak i drugiego – przeznaczonego do analizy ilościowej, co wiązałoby się z koniecznością zakupu mniejszej liczby licencji, a także uniemożliwiało jednoczesną pracę kilku członkom zespołu. Zrezygnowano również z zakupu programu Socjolog 1.5, gdyż udało się zebrać relatywnie mało informacji na temat funkcjonalności tego oprogramowania. Z kolei za wyborem oprogramowania firmy Provalis przemawiała uniwersalność tego narzędzia, która powinna umożliwić przeprowadzenie planowanych analiz jakościowych oraz części analiz ilościowych zgodnie z pierwotnymi założeniami zespołu badawczego. Przede wszystkim QDA

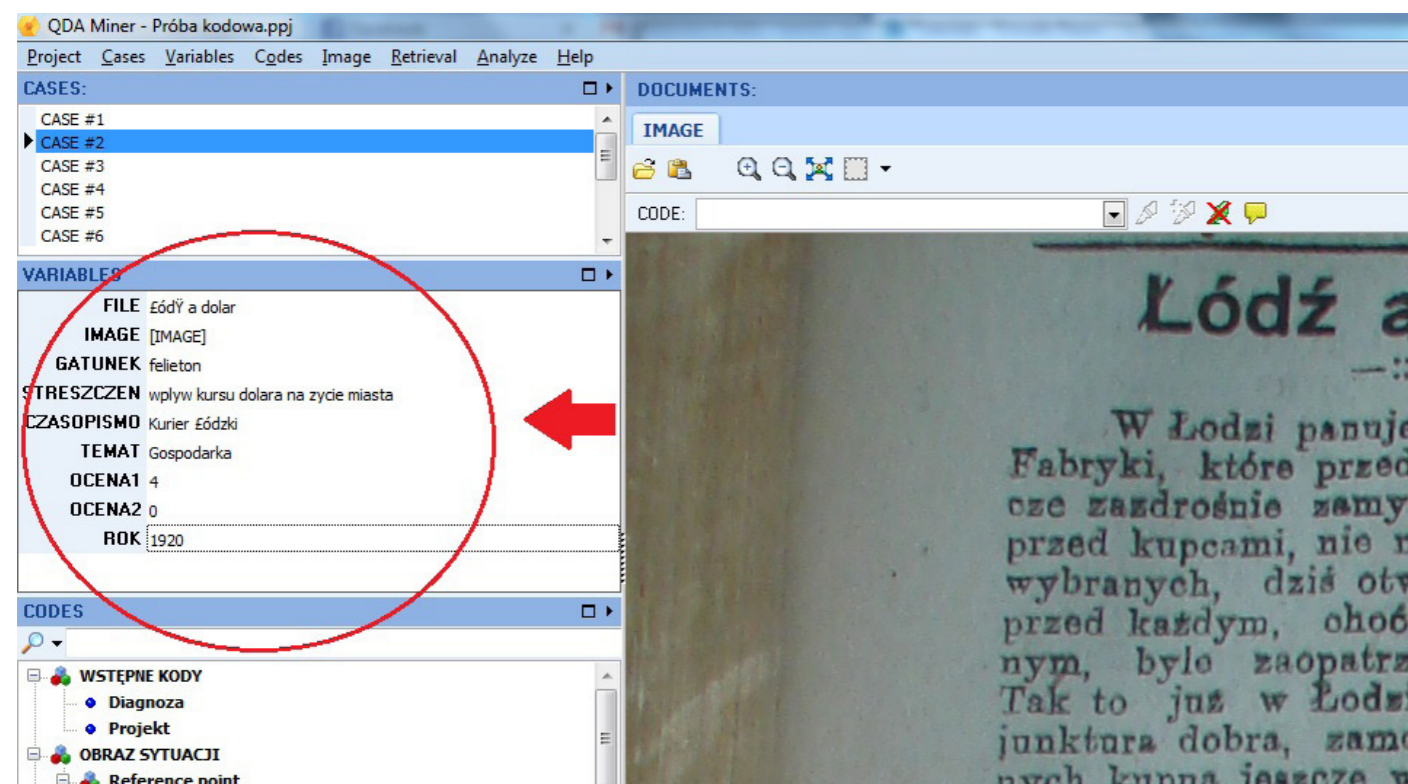
Miner posiada podobne możliwości przeprowadzenia jakościowej analizy materiału empirycznego jak MAXqda, Nvivo czy Atlas.ti. Jednocześnie jego możliwości wraz z dodatkiem w postaci WordStat powinny umożliwić zrealizowanie części badania leksykometrycznego. Ponadto program umożliwia elastyczną i efektywną pracę zespołową, a także daje szerokie możliwości w zakresie scalania dokumentów w jeden projekt, co miało szczególne znaczenie w odniesieniu do liczby członków zespołu badawczego. QDA Miner posiada również moduł (*inter-raters agreement*) służący do oszacowania analizy zgodności, co gwarantuje rzetelne kodowanie wielu kodujących. Kolejnym atutem przemawiającym na korzyść oprogramowania Provalis była jego cena, umożliwiająca nabycie licencji na pięć komputerów, co dało możliwość jednoczesnej pracy znacznej części zespołu nad materiałami z różnych okresów. Podsumowując, w momencie wyboru oprogramowania, po przeprowadzeniu wstępnej weryfikacji dostępnych programów z rodziny CAQDAS, QDA Miner spełniał wszystkie trzy z wyżej przedstawionych warunków oraz wydawał się optymalnym rozwiązaniem na potrzeby projektu.

Wykorzystanie QDA Minera w projekcie

Jak zostało to już wyżej przedstawione, analizie jakościowej i ilościowej poddane zostaną zebrane materiały, głównie w postaci artykułów prasowych, z czterech okresów historycznych. W tym celu zespół projektowy został podzielony na cztery podgrupy, a każda z nich odpowiada za przeprowadzenie kwerendy w jednym z wybranych okresów. Na podstawie przeprowadzonej kwerendy wybranych tytułów z prasy łódzkiej wyselekcjonowane zostaną artykuły,

w których poruszane tematy zgodne są z problematyką projektu. Artykuły te zostaną umieszczone w formie plików graficznych w QDA Minerze. Każda z podgrup odpowiedzialna będzie za przygotowanie bazy artykułów z danego okresu, natomiast w dalszej kolejności bazy te zostaną scalone w jedną ogólną bazę projektową. Przedstawiona procedura stanowi jednocześnie pierwszy etap doboru tekstów do dalszych analiz. Oprócz umieszczenia w bazie każdy z wybranych tekstów zostanie również opisany za pomocą listy zmiennych, które w dalszej kolejności umożliwią łatwe i efektywne zarządzanie zebraniem materiałem empirycznym. W tym miejscu warto odnotować, że opis danego materiału (artykułu, dokumentu urzędowego itp.) za pomocą różnych zmiennych możliwy jest dzięki wykorzystaniu programu QDA Miner

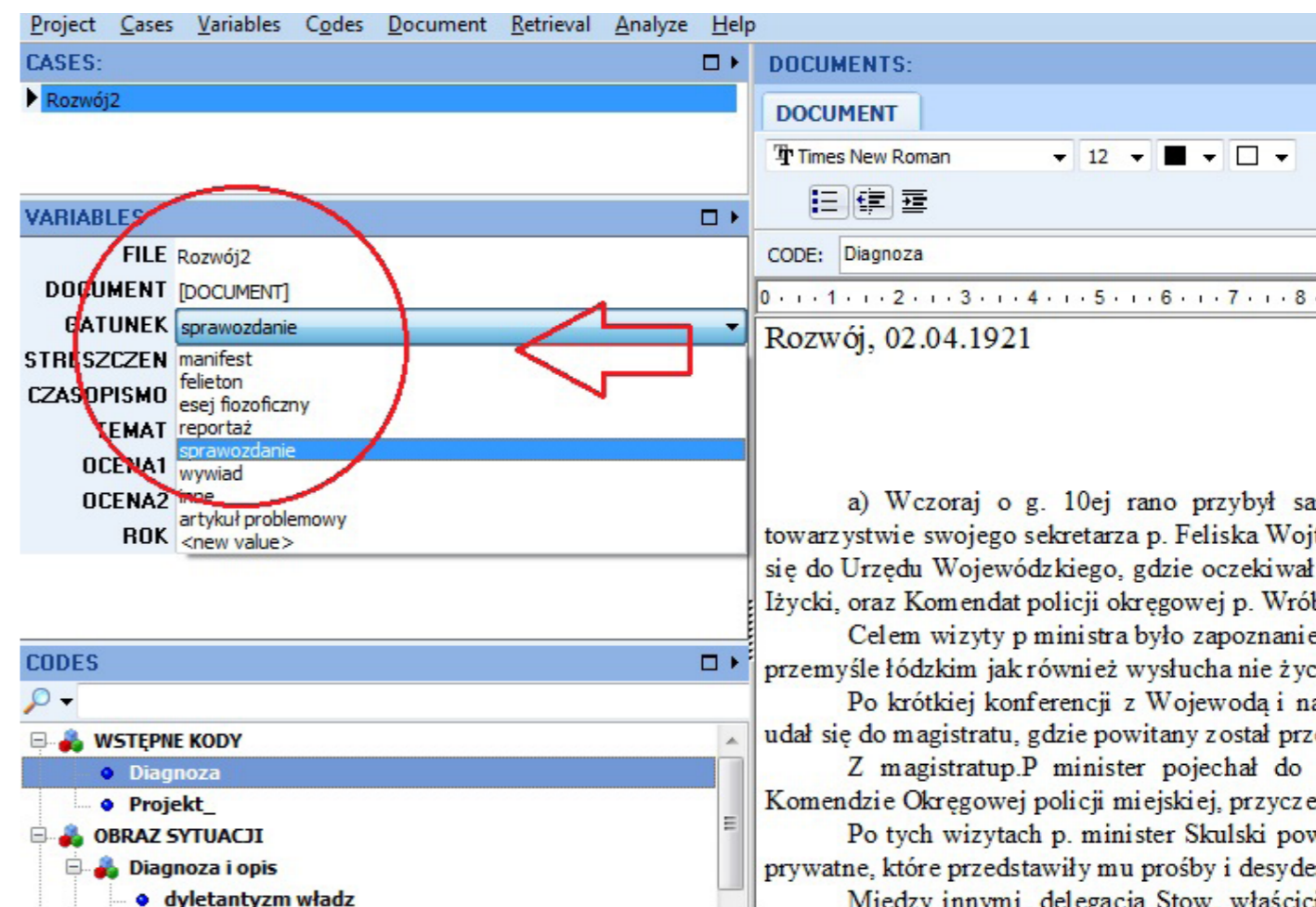
Rys. 1. Zmienne wykorzystywane do opisu tekstu



Źródło: opracowanie własne.

i jego funkcji. Program umożliwia utworzenie wielu zmiennych różnego typu (opisowe, jednokrotnego wyboru, daty itp.) Wśród wypracowanych przez zespół zmiennych znajdują się między innymi: nazwa pliku, gatunek (artykuł, manifest, esej, felieton itp.), krótkie streszczenie, nazwa pisma/dziennika, temat artykułu, a także rok, w którym został opublikowany. Ponadto osoba odpowiedzialna za wybór danego tekstu będzie posiadała możliwość oceny przydatności danego artykułu do dalszej analizy na skali 1–5, na której 1 oznaczać będzie zdecydowanie nieprzydatny, a 5 zdecydowanie przydatny. Teksty, które ocenione zostaną na 4 lub 5, wybrane będą do dalszej analizy. W tym miejscu warto zaznaczyć, że zaproponowana przez badacza ocena powinna zostać zweryfikowana przez jednego z członków podgrupy.

Rys. 2. Przykładowy wybór jednej ze zmiennych opisujących dany tekst

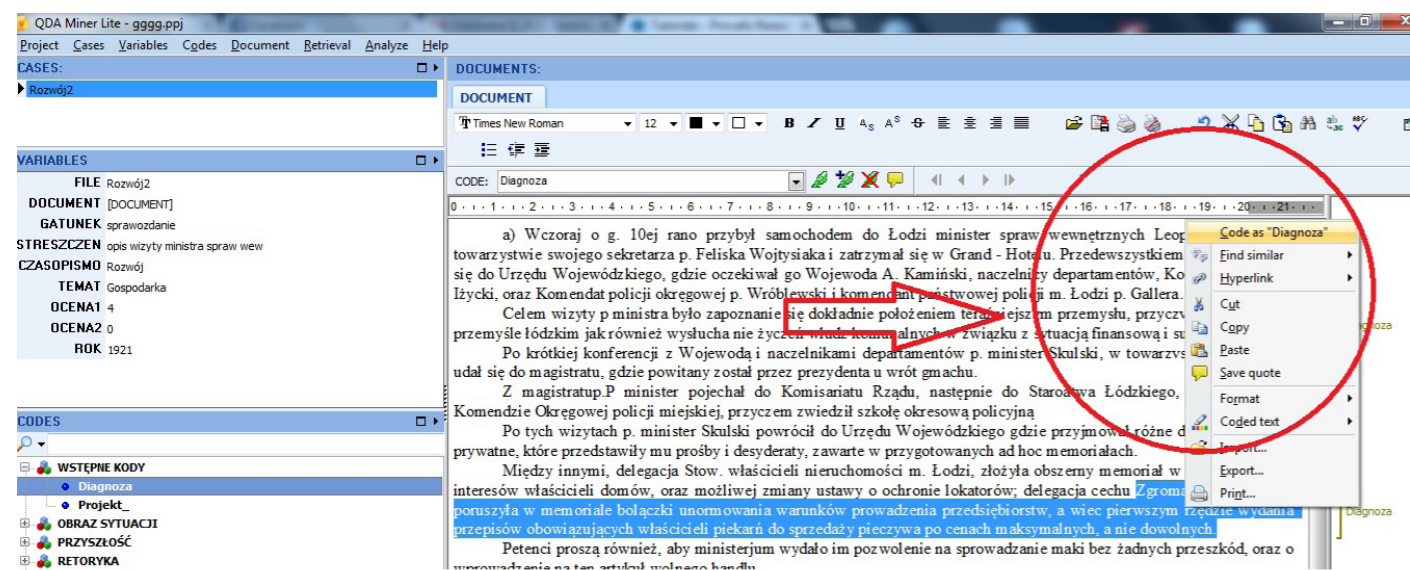


Źródło: opracowanie własne.

Ze względu na bardzo dużą ilość materiału założono, że do dalszych analiz z każdego badanego okresu, na podstawie zaproponowanych ocen, wyselekcjonowanych zostanie około 100 artykułów. Każdy z nich zostanie wstępnie zakodowany za pomocą dwóch najogólniejszych kategorii: „problem” i „diagnoza”, które wynikają z przyjętej ramy teoretycznej. Istotnym elementem jest, żeby zakodowane fragmenty były jak najbardziej syntetyczne i aby streszczały sens przedstawianej „diagnozy” oraz proponowanego „projektu”, który ma przyczynić

się do zmiany zastanej rzeczywistości. W tym miejscu warto zaznaczyć, że kodowanie w programie QDA Miner przebiega podobnie jak w większości programów z rodziny CAQDAS, wystarczy lewym kliknięciem myszy podświetlić jeden z przygotowanych wcześniej kodów, a następnie zaznaczyć wybrany akapit, któremu ma zostać nadany dany kod, kliknąć prawym przyciskiem myszy i wybrać z rozwiniętej listy opcję „code as...”. Jest to istotna cecha, zdecydowanie ułatwiająca i przyspieszająca procesy opracowywania zebranego materiału.

Rys. 3. Przykład nadawania kodów w QDA Minerze



Źródło: opracowanie własne.

Podobnie jak w przypadku pierwszego etapu, badacz odpowiedzialny za zakodowanie danego tekstu będzie również oceniał jego przydatność do dalszych analiz. Ocena również dokonana została za pomocą skali 1–5 i także będzie musiała zostać potwierdzona przez drugiego badacza z podgrupy zajmującej się analizą danego okresu.

W rezultacie ze 100 tekstów zakodowanych za pomocą kategorii „problem” i „diagnoza” planowane jest wyselekcjonowanie z każdego momentu historycznego około 60 tekstów, które poddane zostaną szczegółowemu kodowaniu za pomocą wspólnego dla całego badania klucza, budowanego na podstawie przyjętej ramy teoretycznej. Przed rozpoczęciem procedury wybrane teksty zostaną przekształcone w pliki tekstowe poprzez wykorzystanie techniki OCR. Podstawowe kategorie kodowe zostały wypracowane na podstawie przeglądu literatury przedmiotu oraz seminaryjnych ustaleń zespołu i należy do nich między innymi grupa ko-

dów nazwana roboczo „obrazem sytuacji”. Do niej zaliczają się kolejne kody pojedyncze:

- aktorzy (klasy społeczne, grupy etniczne, grupy religijne, przedstawiciele władzy/administracji);
- diagnoza (problem);
- *reference point* – punkt odniesienia, wzór dla zmiany – zarówno pozytywny, jak i negatywny, zagrożenia itp.;
- obraz stanu przyszłego – po usunięciu przeszkody i bez jej usunięcia;
- nowoczesność (wszelkie odniesienia do postępu, rozwoju, modernizacji; projekt rozwiązania problemu; aktorzy projektu – podmioty i przedmioty);
- przyszłość, a w ramach tej kategorii: porażka, rozwiązanie, podmiot sprawczy oraz przedmiot bierny, obszar interwencji;

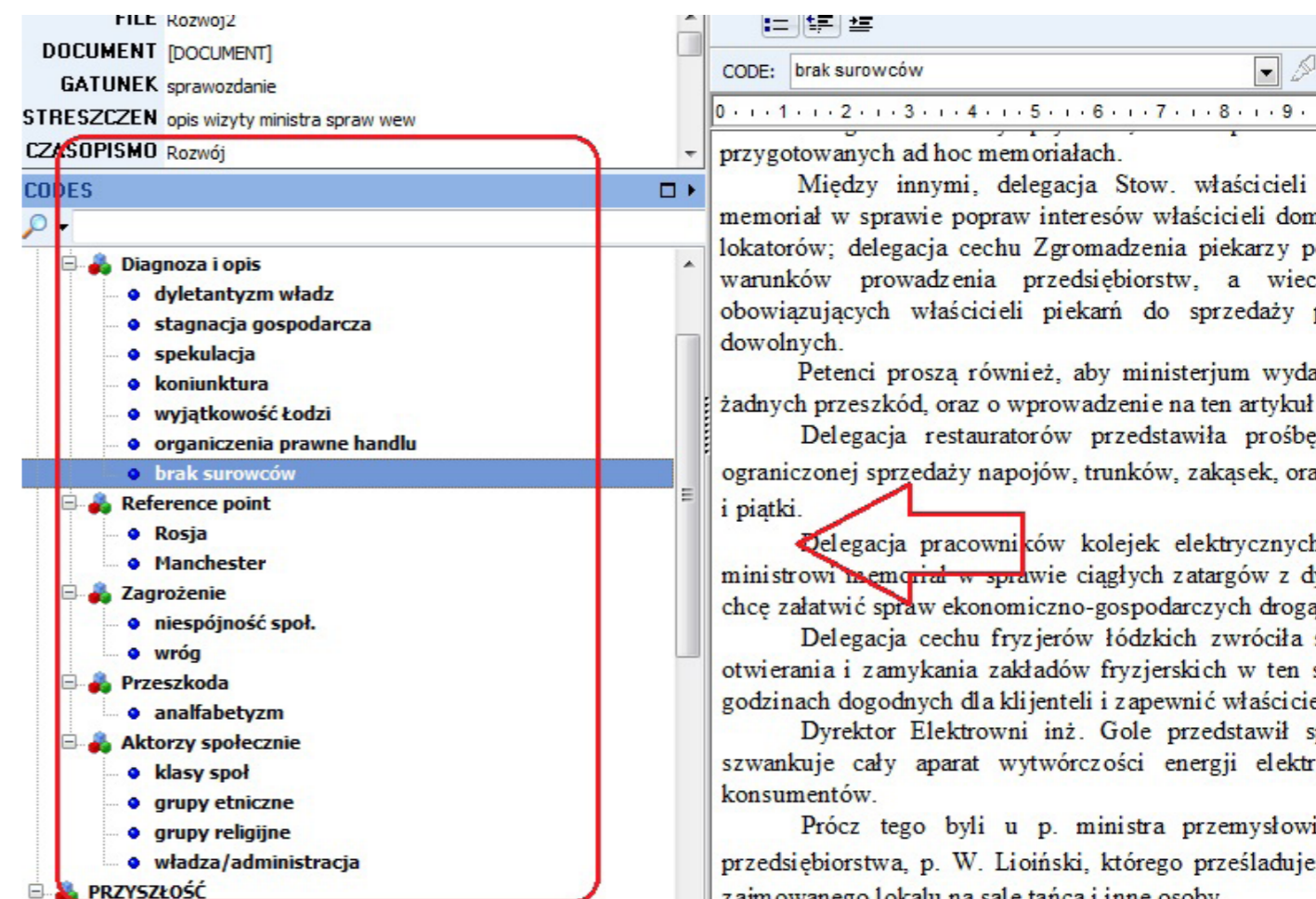
Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

- retoryka (figury retoryczne, konstrukcja argumentu i tym podobne).

W kluczu umieszczono ponadto dodatkowe kategorie, takie jak: topografia (wymiar przestrzenny), Łódź (określenia, opisy) i tło historyczne (istotne wydarzenia, punkty zwrotne). Jednocześnie zaznaczyć należy, że aktualnie trwają prace nad doprecyzowaniem klucza kodowego. Co więcej, mimo że w ramach projektu nie jest planowane wykorzystywanie metodologii teorii ugruntowanej, założono, iż w trakcie kodowania możliwe będzie dodawanie nowych kodów i kategorii, których nie przewidziano w kluczu kodowym.

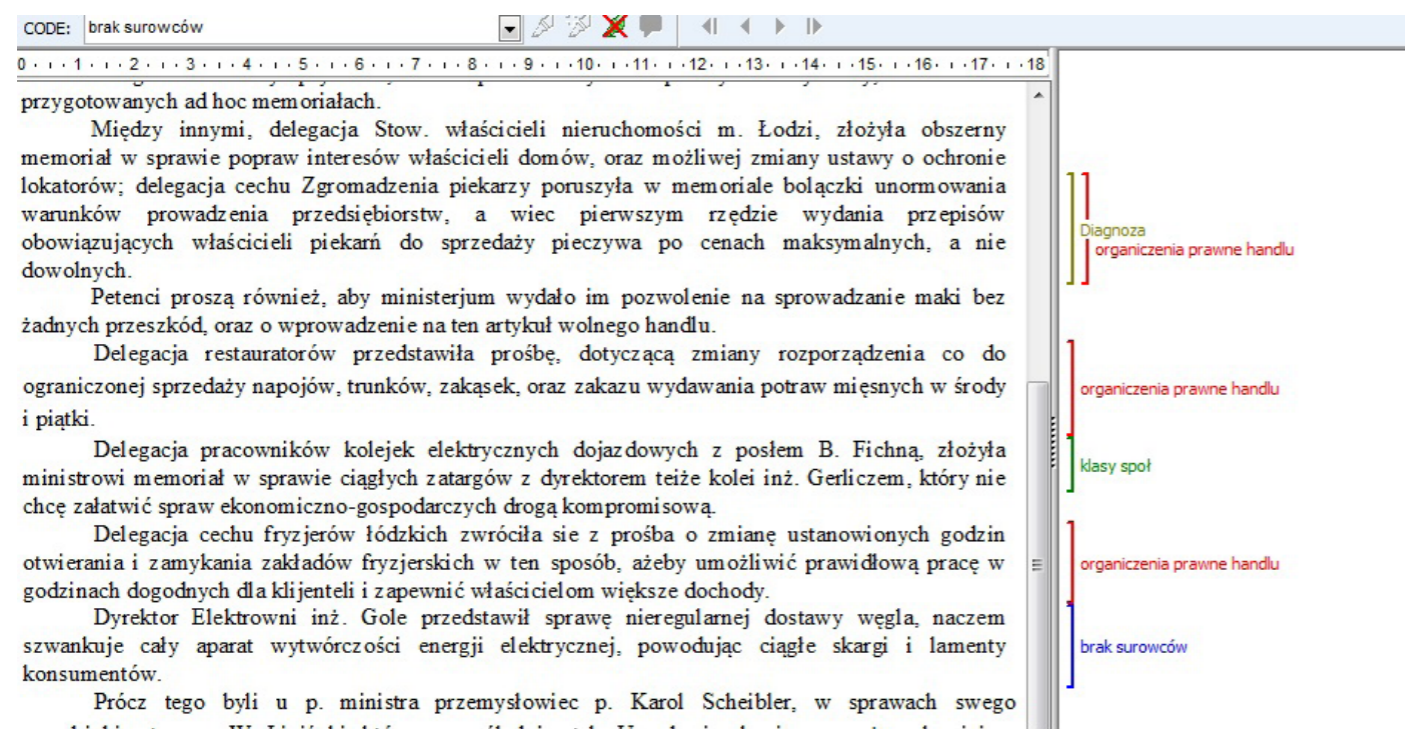
Ze względu na założony plan badawczy możemy jedynie przedstawić opis procedury selekcji materiału i wstępnego kodowania. Przed zespołem badawczym stoi zadanie przeprowadzenia analizy i syntezy jej wyników. Opisany w pierwszej części tekstu cel odnalezienia i opisanie dyskursów nowoczesności jako końcowy efekt pracy nadal jest trudno uchwytne. Na tym etapie możemy jednak podkreślić, że bez komputerowego wsparcia w postaci programu QDA Miner znacznie utrudnione zostałyby działania badawcze. Samo stworzenie dobrze opisanej bazy analizowanych tekstów stanowiłoby znaczny problem w przypadku braku wykorzystania wspomaganie komputerowego.

Rys. 4. Fragment klucza kodowego



Źródło: opracowanie własne.

Rys. 5. Przykładowy fragment zakodowanego artykułu prasowego.



Źródło: opracowanie własne.

Pozytywne i negatywne aspekty wykorzystania programu QDA Miner

Zespół badawczy realizujący opisywany projekt stanął przed trudnym i znaczącym wyzwaniem, jakim było (i nadal jest zgodnie ze stanem w momencie publikacji tego tekstu) podołanie postawionym celom teoretycznym i metodologicznym. Obydwa aspekty projektu były ze sobą ściśle powiązane. Próba odkrycia reguł rządzących dyskursem prasowym w optyce problemu, jakim był sposób zachodzenia przemian modernizacyjnych w Łodzi na przestrzeni wielu dziesiątek lat, nie byłby możliwy bez restrykcyjnego zarządzania pracą zespołu i danymi, które utworzyły analizowany korpus.

Nie ma tu miejsca na szeroki i pogłębiony opis rodzaju danych, które weszły do analizy. Koniecznym jest jednak wskazanie problemu natury metodologicznej i logistycznej, który musiał zostać rozwiązany przed rozpoczęciem etapu zbierania materiału empirycznego. Decyzja o realizacji czterech kwerend obejmujących różne okresy czasu – oddalone od siebie o wiele lat – spowodowała, że konieczne było stworzenie możliwie najbardziej użytecznej platformy do archiwizacji i opisu zebranych tekstów. Co więcej, z racji na założone cele teoretyczne (szczególnie te odnoszące się do leksykometrycznej eksploracji materiału empirycznego), istotne było, by program, który stać się miał wskazaną platformą, umożliwiał wieloaspektową analizę danych.

Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

W toku działań badawczych – wraz z rosnącą liczbą tekstów, które zostały wskazane w toku kwerend jako spełniające podstawowe kryteria kwalifikujące do dalszego procedowania – zespołowi udało się wypracować szczegółowe instrukcje przesiewające zebrany materiał. Celem szczegółowej procedury opisanej w powyższym tekście było ograniczenie korpusu do materiałów nasyconych w możliwie największym stopniu elementami dyskursu, które w ostateczności pozwoliłyby na przeprowadzenie możliwie najbardziej efektywnej analizy tekstu.

Wykorzystanie programu QDA Miner pozwoliło zespołowi na ustrukturyzowanie zebranego korpusu oraz opisanie tekstów zgodnie z założonymi regułami, docelowo kluczem kodowym. Wracając do przywołanych tu pozytywnych aspektów korzystania ze wsparcia komputerowego (Gibbs 2011), możemy podkreślić, że oparcie działań badawczych na tej konkretnej aplikacji zapewniło sprawne stworzenie bazy danych. Jakkolwiek nadal trudno uznać, że proces tworzenia tejże bazy został zakończony – dwie kwerendy nadal są w toku – to dotychczasowe doświadczenia pokazują, iż założony cel połączenia materiałów z różnych epok, opisanie ich, a następnie zakodowania będzie możliwy do osiągnięcia.

Wśród negatywnych stron programu należałoby wymienić ograniczoną stosowalność analiz ilościowych (wspominana wielokrotnie w tekście perspektywa leksykometryczna) z zastrzeżeniem, że jest to bezpośrednio powiązane z dwoma aspektami realizowanego projektu: znacznym zróżnicowaniem językowym analizowanych materiałów oraz ograniczonymi zasobami czasu. Powiązanie tych dwóch czynników uniemożliwiło nam opra-

cowanie zebranego materiału w taki sposób, który pozwoliłby na szersze zastosowanie analiz leksykometrycznych. Brak jednolitego słownika znacznie ograniczył możliwość pracy w kontekście ilościowego ujęcia danych, pozostawiając możliwość aplikacji QDA Miner & Wordstat do analizy zawartości w ramach podziału materiału na cztery kwerendy. Bez powstania słownika dla całego korpusu danych rozszerzenie o bardziej kompleksowe badanie z perspektywy leksykometrycznej jest niemożliwe. Z powodów logistycznych (brak zasobów), ale i merytorycznych (nacisk na analizę jakościową) stworzenie takiego słownika nie zostanie zrealizowane.

W tym sensie trudno mówić o unikalnym charakterze akurat tej aplikacji – wydaje się, opierając swój osąd na opiniach dotyczących innych programów i dotychczasowych doświadczeniach z komputerowym wspomaganem analizy danych, że dla obecnych celów projektu również inne oprogramowania z rodziny CAQDAS spełniłyby podobne warunki. Być może, wraz z postępowaniem prac analitycznych, unikatowość programu nabierze bardziej wyrazistego charakteru, a jego ostateczny wpływ na efekt końcowy projektu będzie znacznie szerszy niż w chwili obecnej.

Podsumowanie

Projekt „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” jest obecnie w połowie realizacji. Przed zespołem badawczym nadal stoją kluczowe działania, które zdecydują o ostatecznym wyniku badań. Dotychczasowe doświadczenia pozwalają nam jednak na postawienie pierwszych ocen wykorzystanego

programu. QDA Miner, połączony z dodatkiem do analiz statystycznych WordStat, zdecydowanie poszerzył zakres możliwości zespołu w odniesieniu do podstawowych działań badacza: opisanie całości zebranego materiału, sklasyfikowania go, przeprowadzenia szczegółowej selekcji i stworzenia struktury danych. Co więcej, proces kodowania analizowanych tekstów przebiega na razie bezproblemowo, a aplikacja sprawdza się jako wygodne narzędzie do zarządzania bazą danych.

Praca zespołowa, której jakość jest kluczowa dla powodzenia całej inicjatywy, została ujęta w szczególne procedury. Decydująca jest tu jednak możliwość szybkiej wymiany informacji i efektywnej współpracy nad konkretnymi materiałami. Jakkolwiek QDA Miner zapewnia wysoką jakość wsparcia dla analizy jakościowej, to trudno odpowiedzieć, czy jego znaczenie jest wyjątkowe na tle innych dostępnych aplikacji (np. pakietu NVivo). Zasadnym byłoby wręcz podkreślenie, że wybór programu w tym względzie nie wskazuje na wyjątkową przewagę czy unikalność propozycji firmy Provalis. Na-

leży podkreślić, że ze względu na powody opisane w tekście, ograniczenie analiz ilościowych i oparcie się w prowadzonych działaniach badawczych głównie na analizach jakościowych dodatkowo zmniejsza rolę aplikacji (szczególnie części WordStat).

Podsumowując, skala projektu bez wątpienia była i jest na tyle duża, że już na etapie przygotowań wymusiła na nas decyzję o połączeniu wysiłków analitycznych z komputerowym wsparciem proponowanym na rynku tego rodzaju usług. Decyzja o wyborze tej konkretnej aplikacji była poprzedzona możliwie najszerszymi konsultacjami. Jakkolwiek fakt napotkanych ograniczeń (brak możliwości pogłębionych analiz leksykometrycznych) wskazuje, że przynajmniej w tym elemencie wybór nie był najbardziej trafny. Nie zmienia to jednak tego, że główny cel, polegający na opisie dyskursów nowoczesności funkcjonujących w różnych momentach historycznych i zależny przede wszystkim od jakościowej interpretacji, jest możliwy również dzięki zastosowaniu opisanego narzędzia komputerowego.

Bibliografia

Flick Uwe (2010) *Projektowanie badania jakościowego*. Przełożył Paweł Tomanek. Warszawa: Wydawnictwo Naukowe PWN.

Gibbs Graham (2011) *Analizowanie danych jakościowych*. Przełożyła Maja Brzozowska-Brywczyńska. Warszawa: Wydawnictwo Naukowe PWN.

Howarth David (2008) *Dyskurs*. Przełożyła Anna Gąsior-Niemiec. Warszawa: Oficyna Naukowa.

Niedbalski Jakub (2012) *OpenCode – narzędzie wspomagające proces przeszukiwania i kodowania danych tekstowych w badaniach jakościowych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 220 [dostęp 25 października 2013 r.]. Dostępny w Internecie: <http://www.qualitativesociologyreview.org/PL/Volume18/PSJ_8_1_Niedbalski.pdf>.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur meto-*

Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

dologii teorii ugruntowanej. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 127–132 [dostęp 25 października 2013 r.]. Dostępny w Internecie: <http://www.qualitativesociologyreview.org/PL/Volume18/PSJ_8_1_Niedbalski_Slezak.pdf>.

Seale Clive (200) *Wykorzystanie komputera w analizie danych jakościowych* [w:] David Silverman, red., *Prowadzenie badań jakościowych*. Przełożyła Joanna Ostrowska. Warszawa: Wydawnictwo Naukowe PWN, s. 232–234.

Cytowanie

Burski Jacek, Brzeziński Kamil (2014) *Próba wykorzystania programu komputerowego QDA Miner do realizacji projektu badawczego „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 82–99 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

The Attempt to Use the Computer Program QDA Miner in the Research Project „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)”

Abstract: The goal of the article is to present the main theoretical and methodological approach of the research project „Cztery dyskursy o nowoczesności – modernizm peryferii na przykładzie Łodzi (XIX–XX wiek)” and to put across the potential support of one of CAQDAS programs for research processes at hand.

In the text, one can find the results of implementation of QDA Miner program (prepared by Provalis company) to the analyses of historical press. Capability and impact the program has on research will be taken under examination. We are going to describe how the characteristics of QDA Miner are impacting upon the construction of empirical database, its selection, categorization, and description. In addition, the article is thought as an extension of previous knowledge about the implementation of computer programs to the process of qualitative analyses.

The structure of the article consists of: brief history of using computer programs in qualitative analyses and short description of research project. In the next part, we present why the QDA Miner was chosen. In the last part of the article, one can find the first impressions and conclusions regarding the use of this program to qualitative analyses of collected material.

Keywords: qualitative analysis, QDA Miner, Lodz, discourse, modernity

Victoria Kamasa Uniwersytet im. Adama Mickiewicza w Poznaniu

DOI: <https://doi.org/10.18778/1733-8069.10.2.06>

Techniki językoznawstwa korpusowego wykorzystywane w krytycznej analizie dyskursu. Przegląd

Abstrakt Krytyczna analiza dyskursu (KAD) jako nurt z pogranicza socjologii i językoznawstwa skupia się na analizie roli dyskursu w umacnianiu i reprodukowaniu relacji władzy i dominacji, zaś badaniu empirycznemu podlegają przede wszystkim różnego rodzaju teksty. Właśnie koncentrację na tekście uznaje się za jedną z przyczyn wzrastającego zainteresowania wykorzystaniem metod językoznawstwa korpusowego w takich analizach. Prezentacji tych metod poświęcony jest niniejszy artykuł. Przedstawione zostaną podstawowe strategie doboru korpusu do badań w ramach KAD oraz najczęściej wykorzystywane metody: analiza frekwencji, analiza słów kluczowych, analiza kolokacji i analiza konkordancji. Zaprezentowane zostaną także przykłady zastosowania poszczególnych metod w różnorodnych tematycznie badaniach. Przegląd metod podsumowany zostanie omówieniem korzyści wynikających z ich zastosowania oraz kosztów, które się z nimi wiążą.

Słowa kluczowe krytyczna analiza dyskursu, metody korpusowe, analiza kolokacji, analiza słów kluczowych, analiza konkordancji

Victoria Kamasa, dr, adiunkt w Instytucie Językoznawstwa na Uniwersytecie im. Adama Mickiewicza w Poznaniu. Jej zainteresowania badawcze mieszczą się na przecięciu tego, co językowe (doktorat z zakresu językoznawstwa stosowanego) i tego, co społeczne (magisterium z socjologii). Obecnie zajmuje się krytyczną analizą dyskursu Kościoła katolickiego w Polsce z wykorzystaniem narzędzi językoznawstwa korpusowego.

Adres kontaktowy:

Instytut Językoznawstwa, Collegium Novum
Uniwersytet im. Adama Mickiewicza w Poznaniu
al. Niepodległości 4, 61-874 Poznań
e-mail: vkamasa@amu.edu.pl

Ogromne zasoby językowe dostępne w sieci, możliwość digitalizacji tekstu, a także jego dalszej cyfrowej obróbki przyniosły znaczące zmiany dla wielu dyscyplin humanistyki i nauk społecznych. I tak, lingwiści zamiast samodzielnie tworzyć lub z trudem gromadzić zaledwie kilka przykładów dla omawianych przez siebie zjawisk, mogą dziś przy pomocy kilku kliknięć uzyskać dostęp do kilkuset, kilku tysięcy, a czasem nawet kilku milionów przykładów (Przepiórkowski i in. 2009). Historycy – dzięki możliwości geograficznego tagowania¹ tek-

¹ Oznaczanie wybranych informacji w tekście umożliwiające następnie automatyczne przeszukiwanie i porządkowanie tekstu wg tych informacji.

stów źródłowych – mogą z niespotykaną dotąd łatwością obserwować przemieszczanie się wojsk czy rozprzestrzenianie się chorób (np. Gregory 2008). Socjologowie mają możliwość korzystania z programów ułatwiających kodowanie analizowanych danych jakościowych, a także późniejsze dostrzeganie wzorców w badanym materiale (np. Schmidt, Skowrońska 2006). Zmiany technologiczne znacząco poszerzyły także możliwości dostępne dla analityków dyskursu, w tym badaczy pracujących w paradygmacie krytycznej analizy dyskursu (KAD). Tym możliwościom, przykładom ich wykorzystania, a także korzyściom i kosztom z nich wynikającym poświęcony jest niniejszy tekst.

Wspomniana powyżej krytyczna analiza dyskursu jest szerokim nurtem badawczym, w którego centrum stoi zainteresowanie relacjami władzy i dominacji, a także rolą dyskursu w umacnianiu i reprodukowaniu tych relacji (van Dijk 2001; Rogers i in. 2005). Wśród celów, jakie stawia sobie KAD, wymienia się między innymi „opis, interpretację i wyjaśnianie relacji między językiem, praktyką społeczną i światem społecznym”² (Rogers i in. 2005: 376), „podnoszenie świadomości dotyczącej strategii używanych, by tworzyć, zachowywać i reprodukować (a)symetryczne relacje władzy, które to strategie realizowane są przy pomocy dyskursu” (Hidalgo Tenorio 2011: 184) czy „identyfikację i interpretację sposobów, w jakie ideologia funkcjonuje w dyskursie i poprzez dyskurs” (Breeze 2011: 520). Z perspektywy niniejszego artykułu szczególnie istotne jest założenie o zasadniczej społecznej roli tekstu będącego manifestacją dyskursu (Wodak

² Wszystkie tłumaczenia pochodzą od autorki.

2001), a także praktyka badawcza, w której analizie poddaje się właśnie teksty, które – choć pochodzą z różnych źródeł (od prywatnych listów, przez artykuły prasowe do debat parlamentarnych) i bywają zapisem wypowiedzi ustnych – łączy to, że są tekstem właśnie.

Koncentracja na tekście jako podstawowym przedmiocie analizy stała się zapewne przyczyną coraz wyraźniejszej tendencji do wykorzystywania narzędzi językoznawstwa korpusowego w analizach KAD (Baker i in. 2008). Sam korpus definiowany jest przeważnie jako zbiór tekstów, które podlegają obróbce maszynowej (ang. *machine-readable texts*) (McEnery, Wilson 2001). Zwraca się także uwagę na jego reprezentatywność i zrównoważenie, przy czym ta pierwsza rozumiana jest jako obecność wszystkich elementów analizowanej odmiany w korpusie (Gries 2009), zaś zrównoważenie jako zachowanie właściwych (uwzględniających częstotliwość i istotność) proporcji pomiędzy reprezentacją poszczególnych elementów badanej odmiany w korpusie (Gries 2009). Baker (2006) z kolei wskazuje na rozróżnienie pomiędzy korpusami ogólnymi (a więc zrównoważonymi ze względu na język jako całość) i korpusami specjalistycznymi, w którym punktem odniesienia dla reprezentatywności jest bądź jakaś odmiana języka (np. język gazet), bądź jakiś konkretny temat (np. uprzejmość). Korpusy zarówno ogólne, jak i specjalistyczne znajdują bardzo szerokie zastosowanie w językoznawstwie i dziedzinach pokrewnych: od badań historii języka (Hebda 2011), przez analizy struktur składniowych i semantycznych (Miechowicz-Mathiasen, Scheffler 2008), aż po badania nad sposobami konceptualizacji wybranych pojęć (Fabiszak, Hebda, Konat 2012).

Są one także coraz częściej używane w analizach prowadzonych w nurcie KAD.

Jak wspomniałam powyżej, celem niniejszego tekstu jest prezentacja podstawowych technik wywodzących się z językoznawstwa korpusowego, które są stosowane w badaniach KAD. Omówię więc poszczególne metody, przedstawiając krótko ich definicje, a także przykłady zastosowania w konkretnych badaniach prowadzonych w ramach KAD. Przegląd badań łączących językoznawstwo korpusowe i KAD nie rości sobie pretensji do zupełności – wybrane przykłady mają raczej ilustrować możliwości analityczne, jakie dają poszczególne techniki, a także wielość i różnorodność tematów, do badania których je wykorzystywano. Przegląd nie ma też na celu prezentacji możliwości, funkcji i ograniczeń konkretnego oprogramowania, a raczej omówienie technik korpusowych, których wykorzystanie stało się możliwe dzięki zwiększeniu mocy obliczeniowych komputerów.

Dobór korpusu

Jedno z podstawowych pytań, jakie nasuwa się w kontekście wykorzystania korpusów w KAD, dotyczy jego wielkości i sposobu doboru. Stosunkowo najrzadziej wykorzystywanym w ramach KAD podejściem jest prowadzenie badań na ogólnych korpusach językowych. I tak na przykład Hamilton i współpracownicy (2007) rekonstruują znaczenie słowa *risk* na podstawie ogólnych korpusów brytyjskiej i amerykańskiej wersji języka angielskiego. Na kwestiach o silniejszym nacechowaniu społecznym skupia swoją uwagę Maunter (2007), analizując funkcjonowanie słowa *elderly* w wybranych podkor-

pusach Bank of English. Z kolei Orpin (2005) porównuje dyskursywne konstrukcje związane z użyciem dwóch angielskich słów określających nieuczciwe zachowania osób publicznych: *corruption* i *sleaze*. Robi to na podstawie analizy ich wystąpień w podkorpusie prasy brytyjskiej korpusu Bank of English. Wspólną cechą opisanych badań jest koncentracja na semantyce badanych słów – szczegółowa analiza kolokacji bądź kontekstów, w których dane słowo występuje, daje bardzo dobry wgląd w jego znaczenie i funkcjonowanie w języku, pozostawia jednak pewien niedosyt w kwestii bardziej społecznie czy krytycznie nastawionych konkluzji.

Zdecydowanie popularniejszym nurtem wśród badaczy KAD jest wykorzystanie specjalistycznych korpusów. Mogą one zostać podzielone według dwóch kryteriów:

1. odmiany języka, którą zawierają: pisane *versus* mówione;
2. części badanej populacji, którą obejmują: próbkowane *versus* pełne.

Najczęściej badaniu podlegają korpusy tekstów pisanych, stanowiące określony wycinek wszystkich tekstów, które można uznać za istotne ze względu na badane zagadnienie. W wielu przypadkach badacze decydują się jednak na tak szczegółowe sformułowanie problemu badawczego, które pozwala im objąć badaniem wszystkie teksty istotne ze względu na postawiony problem.

Tylko nieliczni badacze skupiają się na analizie mowy: badaniu podlegają wtedy transkrypcje nagrań audio pochodzących bądź z naturalnych sy-

tuacji komunikacyjnych (Herbel-Eisenmann, Wagner 2010), bądź z badań prowadzonych metodami nauk społecznych, takimi jak wywiady czy badania fokusowe (Weninger 2010). Do zalet tego typu badań należy koncentracja na najbardziej pierwotnej i spontanicznej formie języka, jaką jest mowa. Można przypuszczać, że o stosunkowo małym wykorzystaniu tego typu korpusów w KAD decydują liczne wyzwania techniczne i organizacyjne związane z ich tworzeniem³.

Z kolei wśród przykładów badań prowadzonych na populacjach znajdziemy analizy oparte na pełnych tekstach debat parlamentarnych dotyczących określonego tematu (Bachmann 2011; Subtirelu 2013), oficjalnych dokumentach wybranych instytucji (Albakry 2004; Freake, Gentil, Sheyholislami 2010; Kamasa 2013a) czy też tekstach medialnych obejmujących wąski okres czasowy wokół wydarzenia uznawanego przez autora za istotne społecznie (Hidalgo Tenorio 2011). Wykorzystanie korpusu obejmującego całą populację tekstów zapewnia wysoki poziom trafności prezentowanych wyników. Z drugiej jednak strony wymusza koncentrację na wąskich zagadnieniach, których istotność ze względu na stawiane sobie przez KAD cele może budzić pewne wątpliwości.

Największą i najbardziej zróżnicowaną grupę analiz stanowią badania oparte na korpusach złożonych z tekstów wybranych w jakiś sposób z szerszego zbioru. Badacze kierują się tu różnorodnymi kryteriami:

³ Do wyzwań tych zaliczyć można między innymi konieczność organizacji nagrań, zapewnienia odpowiedniego poziomu technicznego nagrań czy sporządzenia transkrypcji, która to konieczność wiąże się ze znacznie większymi nakładami środków i czasu niż w przypadku prowadzenia badań na tekstach już zdigitalizowanych, jak na przykład teksty z internetowych wydań gazet.

- intuicyjnym przekonaniem o istotności tekstów wybieranych do korpusu: Salama (2011) analizuje dwie wybrane przez siebie książki, w których prezentowane są odmienne wizje wahhabizmu, a Fortchner i Kolvraa (2012) pracują na korpusie przemówień polityków, których uznają za istotnych aktorów na scenie politycznej;
- wysoką pozycją w wybranych rankingach: Alcaraz-Ariza i Ángeles (2002) tworzą korpus recenzji książek medycznych, wybierając te opublikowane w najlepszych czasopismach z interesującej ją dziedziny, Lischinsky (2011) wykorzystuje raporty największych szwedzkich firm, zaś Prentice (2010) prowadzi analizy na korpusie złożonym z postów opublikowanych na największym forum dotyczącym interesującego ją tematu;
- zawieraniem określonych słów bądź poruszaniem określonych tematów: Koller (2004) pracuje na dwóch subkorpusach złożonych z tekstów opisujących odpowiednio *businessmanów* i *businesswomen*, Lukac (2011) decyduje się na analizę blogów, których autorki otwarcie deklarują bycie *pro-ana*⁴, a Garbielatos i Baker (2008) tworzą złożony algorytm pozwalający im wybrać artykuły prasowe dotyczące imigrantów;
- czasem powstania: Almeida (2011) i Chen (2012) konstruują swoje korpusy, wybierając artykuły opublikowane w określonym przedziale czasowym.

⁴ *Pro-ana* jest nazwą grupy osób uznających anoreksję za styl życia (Lukac 2011).

Wewnątrz zaproponowanych powyżej kategorii stosowane są różnorodne sposoby wyboru konkretnych tekstów, które mają znaleźć się w badanym korpusie. Stosując terminologię opisującą sposób doboru próby w badaniach społecznych (Babbie 2003), możemy tu mówić o doborze losowym (np. Almeida 2011; Chen 2012), celowym (np. Gabrielatos, Baker 2008) czy dostępnościowym (np. Don, Knowles, Fatt 2010). Jak wskazują przytoczone powyżej przykłady, podejście takie umożliwia podejmowanie bardzo różnorodnych zagadnień, tym samym pozwala badaczowi kierować się przy wyborze analizowanego tematu przede wszystkim jego istotnością społeczną (zgodnie z założeniami przyjmowanymi w KAD), nie zaś możliwościami techniczno-organizacyjnymi. W przypadku każdego korpusu powstaje jednak pytanie, na ile wybrane teksty odzwierciedlają dominujące czy też społecznie najistotniejsze dyskursy w danej kwestii.

Korpusy tworzone na potrzeby poszczególnych projektów badawczych zasadniczo różnią się też rozmiarem: zawierają zazwyczaj od kilkudziesięciu (np. O'Halloran 2009) do kilkuset tysięcy słów (np. Bachmann 2011), choć zdarzają się także badania prowadzone na znacznie większych korpusach obejmujących od kilku (np. Don 2010) do nawet kilkuset milionów (np. Gabrielatos, Baker 2008) słów. Wielkość korpusów wydaje się być determinowana zarówno dostępnością tekstów związanych z analizowanym zagadnieniem, jak i technikami, które badacz zamierza wykorzystać: niektóre z nich pozwalają na wyciąganie wniosków jedynie w przypadku bardzo dużych korpusów, podczas gdy inne wręcz przeciwnie – znajdują zastosowanie właściwie wyłącznie w przypadku małych i średnich korpusów.

Na koniec warto zauważyć, że, jak stwierdza Baker (2006: 28), nie ma prostej odpowiedzi dotyczącej tego, jaki rozmiar powinien mieć korpus, a najważniejszym kryterium, którym należy się kierować, podejmując decyzję dotyczącą jego wielkości, jest cel, jakiemu korpus ten ma służyć.

Zastosowanie korpusów w badaniach nad dyskursem wymaga wykorzystania specjalistycznego oprogramowania komputerowego przede wszystkim ze względu na wielkość analizowanego materiału badawczego, a także procedury statystyczne stosowane w poszczególnych technikach. Do najpopularniejszego oprogramowania stosowanego przez badaczy należą bezpłatny AntConc (dostępny na stronie www.antlab.sci.waseda.ac.jp/software.html) i dostępny na licencji WordSmith Tool (Scott 2013b). Oba umożliwiają pracę na korpusach z polskimi znakami diakrytycznymi, obliczanie list frekwencyjnych, list słów kluczowych i list kolokacji dla modyfikowalnych ustawień, a także eksport wyników do wybranych programów. W poszukiwaniu bardziej zaawansowanych narzędzi, warto zwrócić uwagę na te dostępne w ramach Wmatrix, opracowywane przez zespół pod kierownictwem Paula Raysona (2009).

Podstawowe techniki korpusowe stosowane w krytycznej analizie dyskursu

Analiza list frekwencyjnych

Lista frekwencyjna (ang. *word list*, *frequency list*) określana jest jako „lista wszystkich słów pojawiających się w korpusie wraz z częstotliwością ich

występowania i procentowym udziałem w korpusie⁵” (Baker 2006: 51). Jest ona automatycznie generowana przez programy służące do analizy korpusów, takie jak powyżej wspomniane AntConc czy WordSmith Tool. Stanowi również podstawę dla bardziej złożonych narzędzi analitycznych, takich jak na przykład opisane poniżej słowa kluczowe. Sama w sobie może zostać uznana za najbardziej podstawowe narzędzie we wspieranej korpusowo KAD.

Generując listę frekwencyjną, badacz otrzymuje informacje dotyczące wszystkich słów występujących w tekstach poddawanych analizie, jak i częstotliwości ich występowania. Takie dane zostały wykorzystane na przykład przez Chen (2012) w badaniach dotyczących wpływu zmian politycznych w Chinach na prasę. Wpływ ten jest obserwowany przez częstotliwość użycia pozytywnie i negatywnie nacechowanych oraz neutralnych słów porównujących w diachronicznym korpusie artykułów prasowych. Z kolei Mohamad i współpracownicy (2012) wykorzystali listy frekwencyjne stworzone dla dwóch subkorpusów tekstów z podręczników do matematyki do porównania reprezentacji płci w anglojęzycznych podręcznikach wydawanych w Katarze i poza jego granicami. Analiza częstotliwości występowania wybranych zaimków, nazw zawodów czy określeń pokrewieństwa umożliwiła autorom odpowiedź na pytanie dotyczące poziomu seksizmu w obu grupach podręczników. Potencjał listy frekwencyjnej jako narzędzia wspierającego porównywanie różnych dyskursów wykorzystał również Edwards (2012) w badaniach dotyczą-

⁵ Dostępność drugiej z wymienionych informacji jest zależna od używanego oprogramowania.

cych dyskursu Brytyjskiej Partii Narodowej (ang. *British National Party*, *BNP*). Porównanie częstotliwości występowania poszczególnych słów w manifestach BNP z 2005 i 2010 roku umożliwiło mu wybranie słów-węzłów (ang. *node-words*), których częstotliwość istotnie zmieniła się w badanych manifestach i których konteksty występowania zostały następnie poddane dalszej analizie jakościowej. W przypadku analizy list frekwencyjnych oprogramowanie jest więc wykorzystywane w celu ich wygenerowania, podczas gdy wybór jednostek leksykalnych poddawanych dalszej analizie, a także interpretacja uzyskiwanych częstotliwości pozostają po stronie badacza.

Analiza słów kluczowych

Słowa kluczowe określane są jako „punkty, wokół których toczony są bitwy ideologiczne” (Stubbs 2001: 188). Z kolei Mautner (2005), cytując Williamsa, wskazuje na zasadność, a nawet konieczność zainteresowania się słowami kluczowymi, jako że słowa powinny być widziane jako element problemów. Przytoczone wypowiedzi nie dają jednak jasnych wskazówek, jak ustalić, co jest słowem kluczowym dla danego tekstu, dyskursu lub problemu. Spotyka się tu dwa podejścia: pierwsze z nich oparte jest na wiedzy dotyczącej poruszanego zagadnienia, którą posiada badacz. Na jej podstawie wybiera on słowa kluczowe, których występowanie i funkcjonowanie w tekście zostanie później poddane dalszej analizie (np. Mautner 2005; Degano 2007). Drugi sposób poszukiwania słów kluczowych jest bardziej zakorzeniony w tradycji językoznawstwa korpusowego, zgodnie z którą słowo kluczowe definiuje się jako słowo, jakie występuje w jednym korpusie znacząco

częściej niż w innym (określanym jako korpus referencyjny) (Scott 2013a). W celu obliczenia tak zdefiniowanych słów porównuje się za pośrednictwem wspomnianego oprogramowania listę frekwencyjną badanego korpusu z listą frekwencyjną korpusu referencyjnego. Następnie, przy pomocy statystycznych miar kluczowości (np. logarytmiczny wskaźnik wiarygodności [ang. *log-likelihood*] czy test χ^2), generuje się listę słów występujących statystycznie znacząco częściej⁷ w badanym korpusie niż w korpusie referencyjnym.

Przy takiej procedurze określania słów kluczowych szczególnie istotne staje się zagadnienie wyboru korpusu referencyjnego, bowiem to właśnie od jego składu zależeć będzie uzyskany wynik. W tym kontekście Bondi i Scott (2010) zwracają uwagę, że raczej trudno mówić o jednym zakresie tematycznym (ang. *aboutness*) tekstu, o którym informacje pozyskujemy, analizując słowa kluczowe. Za bardziej trafne uznają przyjęcie, że istnieje wiele różnych zakresów tematycznych danego tekstu, które mogą być odkrywane przez zestawienie go z różnymi korpusami referencyjnymi. I tak, chcąc uzyskać ogólne wskazówki dotyczące pojęć istotnych w danym zbiorze tekstów, badacze posługują się zazwyczaj ogólnymi korpusami dla danego języka⁸. Natomiast kiedy ze względu na analizowane zjawisko lub stawiane pytanie badawcze istotny staje się

określony kontekst, wykorzystuje się korpusy referencyjne przygotowywane specjalnie na potrzeby danego badania.

Drugie z opisanych podejść zapewnia koncentrację na wyrażeniach charakterystycznych dla badanego dyskursu, nie zaś na przykład dla gatunku: wykorzystuje je Lukac (2011), porównując badane przez siebie blogi z korpusem blogów na inne tematy. Unika w ten sposób koncentracji na zjawiskach charakterystycznych dla blogu jako gatunku internetowego. Taka strategia okazuje się też bardzo pomocna w ujawnianiu różnic w dyskursach różnych grup na ten sam temat. I tak Subtirelu (2013) wskazuje na różne podejścia do osób nieposługujących się językiem angielskim wśród zwolenników i przeciwników ustawy o wielojęzycznych kartach do głosowania. Pozwala mu je ujawnić analiza słów kluczowych uzyskanych przez zestawienie korpusu wypowiedzi zwolenników ustawy z korpusem wypowiedzi jej przeciwników. Podobną strategię stosuje Baker (2006), zestawiając głosy zwolenników i przeciwników zakazu polowania na lisy z debaty parlamentarnej dotyczącej tego zagadnienia. Dzięki temu ujawnia między innymi dyskursywne powiązanie polowania na lisy z brytyjską tożsamością, obecne w debacie zwolenników polowań czy tendencję do dosłownych i dosadnych określeń efektów polowania wśród jego przeciwników. Z kolei wykorzystanie jako korpusu referencyjnego korpusu ogólnej angielszczyzny FLOB i dalsza analiza uzyskanych w ten sposób słów kluczowych umożliwia Weninger (2010) identyfikację kategorii wspomaganej podmiotowości (ang. *facilitated agency*), charakterystycznej dla amerykańskiego

dyskursu dotyczącego rewitalizacji miast i zrywającej z klasycznym podziałem obrazowania kluczowych uczestników sytuacji społecznych jako pasywnych lub aktywnych.

Wspomniane powyżej przykłady wskazują, że sama lista słów kluczowych nie umożliwia odpowiedzi na pytanie dotyczące dyskursów obecnych w badanym materiale. Konieczna jest dalsza analiza, na którą składa się przyporządkowywanie słów kluczowych do grup tematycznych (np. Gabrielatos, Baker 2008) bądź szczegółowa analiza kontekstów, w których występują w poszukiwaniu wzorców (np. Weninger 2010). Zatem podobnie jak w przypadku list frekwencyjnych program komputerowy generuje jedynie listę słów, zaś jej dalsze opracowanie należy już do badacza. Słowa kluczowe stanowią jednak istotną wskazówkę kierującą uwagę badaczy na zjawiska charakterystyczne dla analizowanego przez nich dyskursu. Mogą być więc swego rodzaju punktem wejścia (ang. *point of entrance*) do zebranych danych.

Analiza kolokacji

Kolokacja jest przez badaczy KAD definiowana zgodnie z tradycją językoznawstwa korpusowego jako częste współwystępowanie (Stubbs 2001). Dla wybranego słowa określa się więc zasięg⁹ (ang. *span*), a następnie przy pomocy oprogramowania wykorzystującego określone miary statystyczne (test t, wskaźnik MI i inne¹⁰) generuje się listę słów występujących istotnie częściej w określonym zasięgu od

⁹ Czyli liczbę słów po prawej i lewej stronie od wybranego słowa, wśród których mają być poszukiwane kolokacje.

¹⁰ Bardziej szczegółowe informacje dotyczące wskaźników statystycznych stosowanych do obliczania kolokacji znaleźć można w Gries (2010).

słowa bazowego. Po uzyskaniu takiej listy, badacze przechodzą do bardziej jakościowej, skupionej na znaczeniu analizy. Spotyka się tu zasadniczo dwa różne modele postępowania:

- koncentracja na samych kolokacjach – słowa z otrzymanej listy łączone są w grupy tematyczne, które pozwalają określić funkcjonowanie kluczowego pojęcia w dyskursie przez odtworzenie najważniejszych domen, z którymi jest kojarzone lub w otoczeniu których występuje. Taką strategię przyjmują na przykład Freaque i współpracownicy (2010) w swoich badaniach dotyczących tożsamości mieszkańców Quebecu. Pozwala im to ustalić na przykład, że dla francuskojęzycznych mieszkańców tej prowincji Quebec wiąże się przede wszystkim z narodem rozumianym jako pewna wspólnota historyczna;
- analiza kontekstu zawierającego otrzymane kolokacje – dla otrzymanych słów generuje się listy ich wystąpień w tekście wraz z najbliższym kontekstem, a następnie analizuje otrzymane fragmenty w poszukiwaniu wzorców. Takie działanie prowadzi Forchtnera i Kolvraa (2012) do ustalenia, że kluczowe dla konstrukcji tożsamości europejskiej w badanym przez nich materiale są wspólnie wyznawane wartości, które są umieszczone w kontekście wspólnej trudnej przeszłości.

Drugim istotnym zagadnieniem związanym z wykorzystaniem analizy kolokacji jest wybór słów, dla których kolokacje te mają zostać ustalone. Również tutaj badacze wybierają różnorodne sposoby

⁶ Więcej informacji na temat miar statystycznych stosowanych dla obliczania słów kluczowych, a także ich ograniczeń można znaleźć np. w Gabrielatos i Marchi (2011; 2012).

⁷ W niektórych badaniach wykorzystuje się także tzw. negatywne słowa kluczowe, czyli takie, które występują w analizowanym korpusie znacząco rzadziej niż w korpusie referencyjnym.

⁸ W przypadku badań dotyczących tekstów w języku polskim możliwe jest skorzystanie z list frekwencyjnych przygotowanych dla Narodowego Korpusu Języka Polskiego (<http://nkjp.uni.lodz.pl/>).

postępowania: Subtirelu (2013) poszukuje kolokacji dla ustalonych wcześniej słów kluczowych. Pozwala mu to skupić uwagę na konstrukcji pojęć charakterystycznych dla badanego przez niego dyskursu¹¹. Z kolei Lischinsky (2011), ze względu na swoje zainteresowanie konstrukcją kryzysu ekonomicznego w prasie, decyduje się skorzystać ze źródeł leksyko-graficznych (słowniki i baza FrameNet) w poszukiwaniu słów używanych, by określać kryzys. Własną intuicją kieruje się z kolei Mautner (2005)¹², określając słowa istotne w jej opinii dla koncepcji *entrepreneurial university* i skupiając swoją uwagę na analizie ich profili kolokacyjnych.

Podsumowując, profile kolokacyjne generowane przy pomocy programu komputerowego wykorzystywane są jako podstawa do identyfikacji konstrukcji dyskursywnych stanowiących przedmiot zainteresowania badacza. Najczęściej służą one uzyskaniu bardziej szczegółowych informacji na temat funkcjonowania określonych słów w badanych tekstach. Na podstawie takiej informacji badacz identyfikuje później struktury dyskursywne. Analiza kolokacji, zwłaszcza prowadzona na dużych korpusach języka ogólnego, pozwala także na ustalenie, jak badane słowo i związany z nim koncept funkcjonują społecznie (np. Mautner 2007; Marling 2010). Lista kolokacji oddaje *stricte* ilościowe relacje pomiędzy słowami występującymi w badanym tekście. Poddana dalszej jakościowej analizie, pozwala jednak zrozumieć złożone relacje pomiędzy reprezentacjami.

¹¹ Podobną strategię stosują również Don, Knowles, Fatt (2010), Freake, Gentil, Sheyholislami (2010) oraz Salama (2011).

¹² Podobną strategię stosują również Mautner (2005; 2007), Hamilton, Adolphs, Nerlich (2007), Forchtner, Kolvraa (2012).

Prozodia semantyczna

Jednym ze szczególnych przypadków wykorzystania list kolokacji jest analiza prozodii semantycznej. Pojęcie to zostało zaproponowane przez Louwa (1993), który zdefiniował ją jako „powtarzającą się konsekwentnie aurę znaczeniową, którą dane słowo zostaje przepojone przez swoje kolokacje” (s. 157). Prozodię semantyczną dla wybranego słowa badacz określa więc na podstawie oceny nacechowania (najczęściej na osi pozytywne–negatywne) jego najsilniejszych kolokacji. Sama koncepcja była poddawana krytyce (zobacz np.: Whitsitt 2005) dotyczącej na przykład możliwości transferu nacechowania z jednej jednostki leksykalnej na inną czy rodzaju nacechowania, jakiego powinny dotyczyć analizy¹³.

Mimo tej krytyki analiza prozodii semantycznej jest wykorzystywana w badaniach z nurtu KAD. I tak na przykład Mautner (2007) wskazuje, że analizowane przez nią słowo *elderly* ma znacznie silniejszą negatywną prozodię semantyczną, kiedy jest używane jako rzeczownik (osoba w podeszłym wieku) niż gdy występuje jako przymiotnik (starszy). Z kolei Kamasa (2013a) pokazuje, jak użycie słowa *praktyka* w odniesieniu do zapłodnienia *in vitro* przyczynia się do jego negatywnej konstrukcji w dyskursie Kościoła katolickiego w Polsce. Analizę prozodii semantycznej wykorzystują również Hamilton, Adolphs i Nerlich (2007) w swoich badaniach nad znaczeniem słów *ryzyko* i *ryzykować* czy Gabrielatos i Baker (2008), którzy analizują dyskursywną konstrukcję emigrantów w brytyjskiej prasie.

¹³ Nacechowanie wyłącznie pozytywne bądź negatywne lub bardziej złożone modele analizy nacechowania (Oster 2010).

Przykłady wykorzystania prozodii semantycznej w KAD wskazują, że jest ona użytecznym narzędziem pozwalającym na identyfikację ukrytych i nieoczywistych wzorców funkcjonowania wybranych pojęć w dyskursie. Z drugiej jednak strony wątpliwości dotyczące realności psychologicznej założenia, że częste występowanie określonego wyrażenia w sąsiedztwie negatywnie nacechowanych wyrażeń powoduje zmianę jego nacechowania, każą zachować daleko posunięta ostrożność w interpretacji wyników takich analiz.

Preferencja semantyczna

Drugim ze sposobów bardziej złożonego wykorzystania list kolokacji w ramach KAD jest analiza preferencji semantycznej. Definiuje się ją jako tendencję określonej jednostki leksykalnej do częstego kolokowania z serią jednostek należących do jednego pola semantycznego (Salama 2011). Podobnie jak w przypadku prozodii semantycznej, preferencja określana jest więc na podstawie listy kolokacji. Z tą różnicą, że tym razem badacz ocenia nie nacechowanie poszczególnych kolokacji, ale ich przynależność do określonych grup tematycznych (pól semantycznych).

Wśród badań, w których analiza preferencji semantycznej prowadzi do identyfikacji społecznych konstrukcji wybranych zjawisk, wskazać można wspomnianą już pracę Hamiltona i współpracowników (2007), w której wykorzystuje się analizę preferencji do identyfikacji dominujących dyskursów związanych z ryzykiem. Prowadzi to na przykład do stwierdzenia obecnej w badanym korpusie tendencji do oceny wielkości ryzyka. Z kolei Sala-

ma (2011) ustala z wykorzystaniem takiej analizy, że sposób reprezentacji wahhabizmu w jednym z badanych przez niego źródeł związany jest z zagrożeniem, konspiracją i koncepcją państwa politycznego.

Koncentracja na grupach tematycznych, do których należą kolokacje słów określających zjawiska, jakimi zajmuje się badacz, prowadzi więc do ujawnienia sposobów jego reprezentowania charakterystycznych dla badanego dyskursu. Za pewną słabość tej analizy można uznać oparcie klasyfikacji do poszczególnych pól semantycznych wyłącznie na intuicji badacza (nie istnieje ani jedna lista pól semantycznych, którą posługują się badacze, ani algorytm pozwalający na przydzielanie słów do takich pól w sposób charakteryzujący się wysokim poziomem intersubiektywności).

Analiza konkordancji

O ile wszystkie opisane powyżej techniki raczej wspierają analizę jakościową bądź dostarczają badaczowi wskazówek, gdzie taką analizę zacząć lub czego może ona dotyczyć, to analiza konkordancji może zostać uznana za najbardziej zbliżoną do klasycznie rozumianej jakościowej analizy dyskursu¹⁴. Samą konkordancję definiuje się jako „listę wszystkich wystąpień poszukiwanego terminu w korpusie, zaprezentowaną wraz z kontekstem, w którym termin ten się pojawia” (Baker 2006: 71). Długość kontekstu jest określana przez badacza i mierzona przy pomocy liczby słów lub znaków (w zależności

¹⁴ Baker i in. (2008) wskazują na przykład, że analiza konkordancji jest jedynym narzędziem językoznawstwa korpusowego, z użyciem którego analitycy dyskursu czują się swobodnie.

od używanego oprogramowania), zaś lista konkordancji jest generowana według wybranych parametrów przez używane oprogramowanie.

Jako że konkordancje dają możliwość obserwowania wybranych terminów w ich najbliższym kontekście i tym samym rekonstrukcji dyskursów związanych z tymi terminami, ich analiza wykorzystywana jest niemal we wszystkich badaniach z nurtu KAD prowadzonych z użyciem narzędzi korpusowych. Przy pomocy powyżej opisanych technik bądź kierując się intuicją, badacz ustala słowa, których kontekst występowania zostaje następnie poddany analizie właśnie na podstawie listy konkordancji. I tak na przykład Albakry (2004), analizując konkordancje w kanadyjskim i amerykańskim raporcie dotyczącym incydentu bratobójczego ognia w Kandaharze w 2002 roku, pokazuje jak grzeczność¹⁵ wpływa na kształt tych raportów. Edwards (2012) w swojej analizie manifestów Brytyjskiej Partii Narodowej, skupiając się na konkordancjach dla słów *our* i *British*, demonstruje wzrastającą tendencję do ukrywania rasizmu i konstrukcji grupy własnej w oparciu o pozornie bardziej inkluzywną kategorię narodowości.

Przykład bardziej osadzonego teoretyczne zastosowania kolokacji znajdziemy u Mulderrig (2011), która koduje badane przez siebie kolokacje według typów akcji (ang. *action-type*), zaproponowanych przez Hallidaya i Matthiessena (2004). Pozwala jej to ustalić diachroniczne zmiany w sposobie reprezentowania rządu w dyskursie laburzystów dotyczącym edukacji w Wielkiej Brytanii. Z kolei Kamasa

¹⁵ Rozumiana w sposób proponowany przez Brown i Levinsona (1987).

(2013b), opierając się na kategoriach teoretycznych zaproponowanych przez van Leeuwena (2008), określa na podstawie odpowiednio wyszukanych konkordancji dyskursywną konstrukcję *rodziny* w oficjalnych dokumentach Kościoła katolickiego w Polsce.

Do zalet analizy konkordancji należy możliwość zbadania kontekstu występowania słów istotnych ze względu na stawiane pytanie badawcze nawet w bardzo dużych korpusach tekstów. Możliwość automatycznego wygenerowania listy wszystkich wystąpień wybranego słowa wraz z jego najbliższym kontekstem znacząco skraca proces analizy, a także podnosi jego stopień trafności (istnieje pewność, że zanalizowane zostały wszystkie wystąpienia danego słowa) i powtarzalności (różni badacze dla tego samego korpusu uzyskują zawsze tę samą listę konkordancji). Poszukiwanie wzorców wśród uzyskanych konkordancji prowadzi badaczy do identyfikacji dominujących dyskursów i sposobów dyskursywnej reprezentacji analizowanych zjawisk. Wadą jest natomiast skupienie się na słowie/słowach, nie zaś problemach, o których w tekście może być mowa, bez użycia wyszukiwanego słowa (np. przy pomocy zaimków).

Ocena zastosowania metod korpusowych w KAD

Wykorzystanie metod korpusowych w badaniach z zakresu KAD wiąże się z istotnymi zmianami w tym polu badawczym. Za najistotniejsze z nich uznajemy znaczące zwiększenie liczby danych poddawanych analizie, podniesienie stopnia przejrzystości stosowanych procedur badawczych oraz kon-

centrację na wzorcach ilościowych. Zmiany te, choć zyskują rosnące grono zwolenników, mogą także budzić pewne obawy i wątpliwości.

Korzyści

Zwiększenie liczby analizowanych danych podnosi trafność uzyskiwanych wyników. Badania oparte na dziesiątkach czy setkach tysięcy słów pochodzących z gazet pozwalają na udzielenie pełniejszej odpowiedzi na pytanie dotyczące reprezentacji na przykład imigrantów niż analiza kilku artykułów. Ponadto możliwość pracy ze stosunkowo dużym korpusem danych daje badaczowi szansę na jego bardziej zrównoważony dobór oraz pozwala zastosować złożone algorytmy wyboru tekstów. Zmniejsza to ryzyko skupienia się na badaniu tekstów, których wyborem kierowała wyłącznie intuicja. Co więcej, jednym z punktów krytyki podnoszonej wobec KAD jest właśnie brak reprezentatywności analizowanych tekstów (Stubbs 1997) czy też kierowanie się osobistymi pobudkami w ich wyborze (Breeze 2011). Zastosowanie dużych korpusów danych stanowi częściową odpowiedź na tę krytykę.

Z kolei podniesienie poziomu przejrzystości stosowanych procedur zwiększa możliwość replikacji prowadzonych badań. Ma również znaczenie w kontekście badań porównawczych: profile kolokacyjne czy słowa kluczowe obliczone przy pomocy określonych metod dla określonych korpusów w jednym języku bądź okresie czasu mogą być porównywane z danymi uzyskanymi w analogiczny sposób dla innego języka bądź innego okresu. Ponadto jawność i przejrzystość metod prowadzących

do uzyskanych wyników mogą zwiększać poziom ich wiarygodności i tym samym czynić je bardziej przekonującymi. Podobnie jak w przypadku analizowanych danych również brak jasnych metod jest jednym z problemów wskazywanych przez krytyków KAD (por. np. Breeze 2011). Zastosowanie wybranych technik korpusowych dostarcza częściowego rozwiązania tego problemu.

Wykorzystanie opartych na wzorcach ilościowych list słów kluczowych czy list kolokacji pozwala badaczom skupić się na najczęściej powtarzających się w tekstach słowach czy konceptach. Uzyskuje się w ten sposób punkt rozpoczęcia dalszych analiz, osadzony w samych danych, nie zaś intuicjach czy wiedzy badacza. Odbiorcy wyników badań zyskują w ten sposób jasność dotyczącą kryteriów wyboru słów, które poddane zostały dalszej analizie, co ułatwia zrozumienie i interpretację prezentowanych rezultatów. Może to również prowadzić do zmniejszenia poziomu stronniczości (ang. *bias*) uzyskiwanych wyników – uwaga badacza zostaje skierowana na kwestie najczęściej pojawiające się w badanych przez niego tekstach, nie zaś na te, które wydają mu się intuicyjnie najistotniejsze.

Uzyskana dzięki oprogramowaniu i technikom korpusowym możliwość obserwacji i interpretacji wzorców ilościowych w badanych tekstach może sprzyjać ujawnianiu ukrytych i nieoczywistych tendencji obecnych w analizowanym materiale. Profile kolokacyjne wybranych słów czy też powtarzalność pewnych sposobów reprezentacji obserwowana przy pomocy analiz konkordancji pozwalają skupić uwagę na regularnościach obecnych w stosunkowo dużym zbiorze danych. Przeniesienie ciężaru

z tego, co powtarzalne w opinii badacza na to, co powtarzalne ze względu na liczbę wystąpień może stanowić również krok w kierunku zwiększenia intersubiektywności prowadzonych analiz.

Wspomniane tutaj korzyści¹⁶ związane z zastosowaniem metod korpusowych w KAD wydają się szczególnie istotne ze względu na przyjmowane w ramach KAD założenie o społecznej roli prowadzonych analiz. Zwiększenie reprezentatywności badanego materiału, przejrzystość stosowanych metod, koncentracja na tym, co najczęstsze czy też obserwacja wzorców ilościowych prowadzą, jak się wydaje, do podniesienia wiarygodności prezentowanych wyników, szczególnie dla szerszego niż tylko specjaliści KAD grona odbiorców. Takie podnoszenie wiarygodności można z kolei uznać za szczególnie istotne ze względu na przyjmowane w ramach KAD założenie o emancypacyjnej roli prowadzonych badań.

Koszty

Założenie o analizie dużych korpusów tekstów może modyfikować zakres podejmowanych tematów badawczych. Kieruje ono bowiem uwagę badacza na teksty, które są łatwo dostępne w wersji elektronicznej i które umożliwiają kompilację stosunkowo dużych zbiorów danych. Tendencję tę widać na przykład w prowadzeniu znaczącej części omówionych tu analiz na tekstach prasowych (np. Koller 2004; Gabrielatos, Baker 2008; Almeida 2011) czy transkrypcjach debat parlamentarnych (np. Bachmann 2011; Subtirelu 2013). Pominięte w ten

¹⁶ Pełniejszy obraz korzyści wynikających z zastosowania technik korpusowych w KAD można znaleźć, sięgając na przykład do Hardt-Mautner (1995), Orpin (2005) czy Bakera (2006).

sposób zostają pytania o dyskursywne konstrukcje obecne w trudniej dostępnych wypowiedziach, takich jak na przykład kazania wpływowych księży, przemówienia liderów wspólnot lokalnych czy wykłady nauczycieli w szkołach i na uniwersytetach. Koncentracja na dużych zasobach tekstowych dostępnych w formie elektronicznej może więc dwojako ograniczać pole badawcze: z jednej strony ogranicza się źródła, na podstawie których szuka się odpowiedzi na pytania badawcze. Z drugiej: ogranicza się także same stawiane pytania do takich, dla których możliwe jest uzyskanie odpowiedzi wyłącznie na podstawie analizy istniejących już i dostępnych elektronicznie tekstów.

Z kolei skupienie się na wzorcach ilościowych każe zadać pytanie, czy najczęstszy oznacza najważniejszy. Wydaje się, że brak jest przekonujących dowodów, że to właśnie częstość występowania w dyskursie stanowi czynnik najsilniej wpływający na kształt społecznych i indywidualnych reprezentacji wybranego zjawiska. Pominięta zostaje na przykład kwestia zróżnicowanej społecznie definiowanej istotności tekstów umieszczanych w korpusie. Można przypuszczać, że pewne wypowiedzi, czy to ze względu na ich autorów, czy też kształt samej wypowiedzi (np. szczególną wyrazistość), mogą mieć silniejszy wpływ na konstrukcje dyskursywne obecne wśród odbiorców.

Na uwagę zasługuje także problem wielu zmiennych, które wpływają na uzyskiwane przy pomocy technik korpusowych wyniki. Listy słów kluczowych zależą od trafności doboru zarówno głównego korpusu, jak i korpusu referencyjnego, a także od dostosowania stosowanej miary statystycznej.

Uzyskiwane profile kolokacyjne są również zależne od współczynników, przy pomocy których są obliczane, a ich wiarygodność jako źródła informacji dotyczącego konstrukcji dyskursywnych zależy także od trafności wyboru słów, dla których są obliczane. Drobny błąd czy nieścisłość pojawiająca się na którymkolwiek etapie wykorzystania technik korpusowych mogą prowadzić do skrzywienia wyników, które trudno będzie dostrzec zarówno samemu badaczowi, jak i odbiorcom jego badań.

Innym problemem, który związany jest również z samymi technikami korpusowymi, jest występowanie w nich licznych, niewyrażanych wprost i często nieuwzględnianych w interpretacjach wyników założeń. I tak na przykład wykorzystanie techniki słów kluczowych prowadzi do koncentracji na tym, co różne i pominięcia tego, co podobne. Analizie poddane zostają słowa, które najbardziej różnią się częstotliwością w badanych korpusach. Pomija się na przykład te, które występują bardzo często zarówno w korpusie głównym, jak i referencyjnym. Analiza konkordancji dla określonych słów związana jest z rekonstrukcją funkcjonowania w dyskursie wybranego słowa, nie zaś konceptu, jaki reprezentuje, a który może być wrażany także przy pomocy innych słów. Dodatkowo prowadzi ona do uznania najbliższego kontekstu danego słowa za najistotniejszy i decydujący o wynikach prowadzonych badań. Wybór kolokacji jako narzędzia ułatwiającego badanie dyskursu zakłada, że istotna jest tylko częstość współwystępowania, nie zaś inne cechy połączenia między wyrazami. Każde z tych założeń może zostać podane w wątpliwość. Każde z nich wpływa także na zakres i sposób interpretacji uzyskiwanych wyników.

Na koniec warto także wspomnieć o kwestii specyficznej dla analiz prowadzonych na teksach w języku polskim. Opisane powyżej techniki wykorzystywane są przede wszystkim dla danych w języku angielskim, który charakteryzuje się inną strukturą morfosyntaktyczną niż polski: brak odmiany rzeczownika, a także występowanie zaledwie kilku form czasownika ma wpływ na wyniki ilościowych porównań prowadzonych dla słów rozumianych jako określone sekwencje znaków. To samo pojęcie¹⁷ w języku angielskim wyrażane jest przy pomocy zaledwie kilku różnych słów, w polskim zaś kilkunastu lub kilkadziesiątu, co osłabia analityczną siłę narzędzi, takich jak listy kolokacji czy słów kluczowych, w przypadku badań prowadzonych na niezlematyzowanych¹⁸ korpusach. Z kolei lematyzacja pociąga za sobą liczne problemy, jak na przykład możliwości przypisania określonej formy tekstowej do różnych lematów, która z kolei może prowadzić od zniekształcenia oryginalnego tekstu¹⁹.

Większość wspomnianych tu problemów może zostać rozwiązana przez przemyślaną i precyzyjną konstrukcję korpusów, a także uwzględnienie możliwych wątpliwości w opracowaniu i interpretacji wyników. Ich pominięcie może jednak prowadzić do niebezpiecznej sytuacji: uzyskiwane przy pomocy metod korpusowych wyniki stwarzają wrażenie bardzo wiarygodnych ze względu na ilość danych, na których są oparte, a także ograniczenie roli badacza-

¹⁷ Przyjmujemy tutaj uproszczony model, w którym jednemu pojęciu odpowiada jeden leksem.

¹⁸ Lematyzacja jest to proces przypisania każdej formie wyrazowej występującej w tekście jej formy podstawowej (lematu).

¹⁹ Bardziej szczegółowe informacje na temat lingwistyczno-technicznych problemów związanych z przetwarzaniem polszczyzny można znaleźć na przykład w Młodzki, Przepiórkowski (2009) czy Głowińska, Przepiórkowski (2010).

-interpretatora na rzecz wykorzystania miar statystycznych. Jednak przez niedociągnięcia w procesie analizowania i interpretowania danych, mogą mieć *de facto* niską wartość poznawczą.

Podsumowanie

Zaprezentowane powyżej metody korpusowe stanowią przegląd technik najczęściej stosowanych w badaniach prowadzonych w ramach KAD. Opisane przykłady badań obrazują wielość pytań badawczych, na które szuka się odpowiedzi przy ich pomocy, a także różnorodność tekstów, do których mogą być stosowane. Przedstawione korzyści mogą zachęcać do stosowania technik korpusowych, zaś koszty wskazywać punkty, w których należy zachować szczególną ostrożność przy projektowaniu badań i interpretacji ich wyników.

Bibliografia

Albakry Mohammed (2004) *U.S. "Friendly Fire" Bombing of Canadian Troops: Analysis of the Investigative Reports*. „Critical Inquiry in Language Studies”, vol. 1, no. 3, s. 163–178.

Alcaraz-Ariza, María Ángeles (2002) *Evaluation in English-Medium Medical Book Reviews*. „International Journal of English Studies”, vol. 2, no. 1, s. 137–153.

Almeida Eugenie P. (2011) *Palestinian and Israeli Voices in Five Years of U.S. Newspaper Discourse*. „International Journal of Communication”, vol. 5, s. 1586–1605.

Babbie Earl R. (2003) *Badania społeczne w praktyce*. Przełożyła Agnieszka Kloskowska-Dudzińska. Warszawa: Wydawnictwo Naukowe PWN.

Zaproponowany tu przegląd nie ma w żadnym wymiarze wyczerpującego charakteru. W ramach wspieranej korpusowo KAD stosuje się także na przykład analizę wiązek leksykalnych (Herbel-Eisenmann, Wagner 2010) czy automatyczne tagowanie semantyczne wraz z wykorzystaniem opisanych technik dla kategorii semantycznych, nie zaś poszczególnych słów (Prentice 2010). Metody korpusowe stosuje się także do operacjonalizacji kategorii analitycznych obecnych w różnych szkołach w ramach KAD, takich jak krytyczno-historyczna analiza dyskursu Wodak (O'Halloran 2009) czy propozycje van Leeuwena (Kamasa 2013b; Subtirelu 2013). Lista kosztów i korzyści zależy zaś od przyjmowanych założeń filozoficznych i teoretycznych, perspektywy, z której mają być prowadzone badania, a także przyzwyczajień i przekonań samego badacza.

Bachmann Ingo (2011) *Civil partnership – "gay marriage in all but name": a corpus-driven analysis of discourses of same-sex relationships in the UK Parliament*. „Corpora”, vol. 6, no. 1, s. 77–105.

Baker Paul (2006) *Using corpora in discourse analysis*. London, New York: Continuum.

Baker Paul i in. (2008) *A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK Press*. „Discourse & Society”, vol. 19, no. 3, s. 273–306.

Bondi Marina, Scott Mike (2010) *Keyness in texts*. Amsterdam, Philadelphia: John Benjamins Publishing.

Breeze Ruth (2011) *Critical Discourse Analysis and Its Critics*. „Pragmatics”, vol. 21, no. 4, s. 493–525.

Brown Penelope, Levison Stephen C. (1987) *Politeness. Some universals in language usage*. Cambridge: Cambridge University Press.

Chen Lily (2012) *Reporting news in China: Evaluation as an indicator of change in the China Daily*. „China Information”, vol. 26, no. 3, s. 303–329.

Degano Chiara (2007) *Dissociation and Presupposition in Discourse: A Corpus Study*. „Argumentation”, vol. 21, no. 4, s. 361–378.

Don Zuraidah Mohd, Knowles Gerry, Fatt Choong Kwai (2010) *Nationhood and Malaysian identity: a corpus-based approach*. „Text & Talk – An Interdisciplinary Journal of Language, Discourse & Communication Studies”, vol. 30, no. 3, s. 267–287.

Edwards Geraint O. (2012) *A comparative discourse analysis of the construction of 'in-groups' in the 2005 and 2010 manifestos of the British National Party*. „Discourse & Society”, vol. 23, no. 3, s. 245–258.

Fabiszak Małgorzata, Hebda Anna, Konat Barbara (2012) *Dichotomy between private and public experience: The case of Polish wierzyć 'believe'*. „Selected Papers from UK-CLA Meetings”, vol. 1, s. 164–176.

Forchtner Bernhard, Kolvraa Christoffer (2012) *Narrating a 'new Europe': From 'bitter past' to self-righteousness?* „Discourse & Society”, vol. 23, no. 4, s. 377–400.

Freake Rachele, Gentil Guillaume, Sheyholislami Jaffer (2010) *A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec*. „Discourse & Society”, vol. 22, no. 1, s. 21–47.

Gabrielatos Costas, Baker Paul (2008) *Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005*. „Journal of English Linguistics”, vol. 36, no. 1, s. 5–38.

Gabrielatos Costas, Marchi Anna (2011) *Keyness: Matching metrics to definitions*. Refereat zaprezentowany podczas *Theoretical-methodological challenges in corpus approaches to discourse studies - and some ways of addressing them*, 5 listopada, Portsmouth, Anglia.

Gabrielatos Costas, Marchi Anna (2012) *Keyness: Appropriate metrics and practical issues*. Refereat zaprezentowany podczas *CADS International Conference 2012*, 13–14 września, Bologna, Włochy.

Głowińska Katarzyna, Przepiórkowski Adam (2010) *The Design of Syntactic Annotation Levels in the National Corpus of Polish* [dostęp 15 stycznia 2014 r.]. Dostępny w Internecie: <http://nlp.ipipan.waw.pl/~adamp/Papers/2010-Irec-kg/Irec-nkjp_0324.pdf>.

Gregory Ian N. (2008) *Different Places, Different Stories: Infant Mortality Decline in England and Wales, 1851–1911*. „Annals of the Association of American Geographers”, vol. 98, no. 4, s. 773–794.

Gries Stefan Thomas (2009) *Quantitative corpus linguistics with R. A practical introduction*. New York: Routledge.

Gries Stefan Thomas (2010) *Useful statistics for corpus linguistics* [w:] Aquilino Sánchez, Moisés Almela, eds., *A mosaic of corpus linguistics: selected approaches*. Frankfurt am Main: Peter Lang, s. 269–291.

Halliday Michael, Matthiessen Christian (2004) *An introduction to functional grammar*. London, New York: Arnold.

Hamilton Craig, Adolphs Svenja, Nerlich Brigitte (2007) *The meanings of 'risk': a view from corpus linguistics*. „Discourse & Society”, vol. 18, no. 2, s. 163–181.

Hardt-Mautner Gerlinde (1995) *'Only Connect.' Critical Discourse Analysis and Corpus Linguistics* [dostęp: 20 maja 2012 r.]. Dostępny w Internecie: <<http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf>>.

Hebda Anna (2011) *Onde and envy: A diachronic cognitive approach* [w:] Jacek Fisiak, ed., *Studies in Old and Middle English*. Frankfurt am Main: Peter Lang, s. 107–126.

Herbel-Eisenmann Beth, Wagner David (2010) *Appraising lexical bundles in mathematics classroom discourse: obligation and choice*. „Educ Stud Math”, vol. 75, no. 1, s. 43–63.

Hidalgo Tenorio Encarnacion (2011) *Critical Discourse Analysis, An overview*. „Nordic Journal of English Studies”, vol. 10, no. 1, s. 184–210.

Kamasa Victoria (2013a) *Naming "In Vitro Fertilization": Critical Discourse Analysis of the Polish Catholic Church's Official Do-*

cuments. „Procedia – Social and Behavioral Sciences”, vol. 95, s. 154–159.

Kamasa Victoria (2013b) *Rodzina w dyskursie Polskiego Kościoła Katolickiego. Badania korpusowe z perspektywy Krytycznej Analizy Dyskursu*. „Socjolingwistyka”, vol. 27, s. 139–152.

Koller Veronika (2004) *Businesswomen and war metaphors: ‘Possessive, jealous and pugnacious’?* „Journal of Sociolinguistics”, vol. 8, no. 1, s. 3–22.

Lischinsky Alon (2011) *In times of crisis: a corpus approach to the construction of the global financial crisis in annual reports*. „Critical Discourse Studies”, vol. 8, no. 3, s. 153–168.

Louw Bill (1993) *Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies* [w:] M. Baker, ed., *Text and Technology*. Amsterdam: John Benjamins, s. 157–176.

Lukac Morana (2011) *Down to the bone: A corpus-based critical discourse analysis of pro-eating disorder blogs*. „Jezikoslovje”, vol. 12.2, s. 187–209.

Marling Raili (2010) *The Intimidating Other: Feminist Critical Discourse Analysis of the Representation of Feminism in Estonian Print Media*. „NORA – Nordic Journal of Feminist and Gender Research”, vol. 18, no. 1, s. 7–19.

Mautner Gerlinde (2005) *The Entrepreneurial University*. „Critical Discourse Studies”, vol. 2, no. 2, s. 95–120.

----- (2007) *Mining large corpora for social information: The case of elderly*. „Language in Society”, vol. 36, no. 1, s. 51–72.

McEnery Tony, Wilson Andrew (2001) *Corpus linguistics. An introduction*. Edinburgh: Edinburgh University Press.

Miechowicz-Mathiasen Katarzyna, Scheffler Paweł (2008) *A corpus-based analysis of the Polish verb podobać się* [w:] Jacek Witkoś, ed., *Elements of Slavic and Germanic grammars: a comparative view. Papers on topical issues in syntax and morphosyntax*. Frankfurt, Berlin, Bern, Brussels, New York, Oxford, Vienna: Peter Lang, s. 89–111.

Młodzki Rafał, Przepiórkowski Adam (2009) *The WSD Development Environment* [w:] Zygmunt Vetulani, ed., *Proceedings*

of LTC 2009, s. 185–189 [dostęp 15 stycznia 2014 r.]. Dostępny w Internecie: <<http://nlp.ipipan.waw.pl/~adamp/Papers/2009-ltc-wsd/ltc-034-mlodzki.pdf>>.

Mohamad Subakir Mohd Yasin i in. (2012) *Linguistic Sexism In Qatari Primary Mathematics Textbooks*. „GEMA Online™ Journal of Language Studies”, vol. 12, no. 1, s. 53–68.

Mulderrig Jane (2011) *Manufacturing Consent: A corpus-based critical discourse analysis of New Labour’s educational governance*. „Educational Philosophy and Theory”, vol. 43, no. 6, s. 562–578.

O’Halloran Kieran (2009) *Inferencing and cultural reproduction: a corpus-based critical discourse analysis*. „Text & Talk – An Interdisciplinary Journal of Language, Discourse Communication Studies”, vol. 29, no. 1, s. 21–51.

Orpin Debbie (2005) *Corpus Linguistics and Critical Discourse Analysis. Examining the ideology of sleaze*. „International Journal of Corpus Linguistics”, vol. 10:1, s. 37–61.

Oster Ulrike (2010) *Using corpus methodology for semantic and pragmatic analyses: What can corpora tell us about the linguistic expression of emotions?* „Cognitive Linguistics”, vol. 21, no. 4, s. 727–763.

Prentice Sheryl (2010) *Using automated semantic tagging in Critical Discourse Analysis: A case study on Scottish independence from a Scottish nationalist perspective*. „Discourse & Society”, vol. 21, no. 4, s. 405–437.

Przepiórkowski Adam i in. (2009) *Narodowy Korpus Języka Polskiego*. „Biuletyn Polskiego Towarzystwa Językoznawczego”, t. 65, s. 47–56.

Rayson Paul (2009) *Wmatrix: a web-based corpus processing environment*, Computing Department, Lancaster University, dostępny na stronie <<http://ucrel.lancs.ac.uk/wmatrix>>.

Rogers Rebecca i in. (2005) *Critical Discourse Analysis in Education: A Review of the Literature*. „Review of Educational Research”, vol. 75, no. 3, s. 365–416.

Salama Amir H.Y. (2011) *Ideological collocation and the recontextualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis*. „Discourse & Society”, vol. 22, no. 3, s. 315–342.

Schmidt Filip, Skowrońska Marta (2006) *Człowiek w sieci przedmiotów. Socjologiczna analiza roli i znaczenia przedmiotów w przestrzeni domowej* [w:] Jacek Kowalewski, Wojciech Piasek, Marek Śliwa, red., *Rzeczy i ludzie. Humanistyka wobec materialności*. Olsztyn: Colloquia Humaniorum, s. 197–222.

Scott Mike (2013a) *WordSmith Tools Help* [dostęp 20 sierpnia 2013 r.]. Dostępny w Internecie: <http://www.lexically.net/downloads/version6/HTML/index.html?keywords_info.htm>.

----- (2013b) *WordSmith Tools*. Liverpool: Lexical Analysis Software.

Stubbs Michael (1997) *Whorf’s Children: Critical comments on Critical Discourse Analysis (CDA)* [w:] Ann Ryan, Alison Wray, eds., *Evolving models of language. Papers from the annual meeting of the British Association for Applied Linguistics held at the University of Wales, Swansea, September 1996*. Clevedon: British Association for Applied Linguistics, s. 100–116.

----- (2001) *Words and phrases. Corpus studies of lexical semantics*. Oxford, Malden: Blackwell Publishers.

Cytowanie

Kamasa Victoria (2014) *Techniki językoznawstwa korpusowego wykorzystywane w krytycznej analizie dyskursu. Przegląd*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 100–117 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakoosciowej.org>.

Corpus Linguistics Techniques Used for Critical Discourse Analysis. An overview

Abstract: The paper aims to present corpus methods most commonly used in Critical Discourse Analysis (CDA). The issues of corpus design for CDA will be discussed and methods frequently used in such analysis will be presented: frequency lists, keywords, collocations, and concordances. Moreover, examples of research using this methods will be overviewed to provide an account of the variety of subjects and conclusions the discussed methods might lead to. The paper will conclude with some remarks on benefits and costs related to the usage of corpus methods in CDA.

Keywords: Critical Discourse Analysis, corpus methods, collocation, key words, collocations, concordances

Krzysztof Tomanek
Uniwersytet Jagielloński

DOI: <https://doi.org/10.18778/1733-8069.10.2.07>

Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych

Abstrakt Celem artykułu jest prezentacja podstawowych metod klasyfikacji jakościowych danych tekstowych. Metody te korzystają z osiągnięć wypracowanych w takich obszarach, jak przetwarzanie języka naturalnego i analiza danych nieustrukturalizowanych. Przedstawiam i porównuję dwie techniki analityczne stosowane wobec danych tekstowych. Pierwsza to analiza z zastosowaniem słownika tematycznego. Druga technika oparta jest na idei klasyfikacji Bayesa i opiera się na rozwiązaniu zwanym naiwnym klasyfikatorem Bayesa. Porównuję efektywność dwóch wspomnianych technik analitycznych w ramach analizy sentymentu. Akcentuję rozwiązania mające na celu zbudowanie trafnego, w kontekście klasyfikacji tekstów, słownika. Porównuję skuteczność tak zwanych analiz nadzorowanych do skuteczności analiz zautomatyzowanych. Wyniki, które prezentuję, wzmacniają wniosek, którego treść brzmi: słownik, który w przeszłości uzyskał dobrą ocenę jako narzędzie klasyfikacyjne, gdy stosowany jest wobec nowego materiału empirycznego, powinien przejść fazę ewaluacji. Jest to, w proponowanym przeze mnie podejściu, podstawowy proces adaptacji słownika analitycznego, traktowanego jako narzędzie klasyfikacji tekstów.

Słowa kluczowe analiza danych jakościowych, analiza sentymentu, analiza treści, text mining, kodowanie tekstów, przetwarzanie języka naturalnego, słownik RID, naiwny klasyfikator Bayesa, CAQDAS

Krzysztof Tomanek, doktorant w Instytucie Socjologii Uniwersytetu Jagiellońskiego. Jego zainteresowania badawcze dotyczą zagadnień lojalności, teorii zaufania, zagadnienia Quality of Life w badaniach społecznych. Najważniejsze zainteresowania metodologiczne obejmują zastosowanie technik text mining do analiz danych jakościowych, analizy danych jakościowych wspierane rozwiązaniami NLP, SVR. Prowadzi grant badawczy MNiSW dotyczący Festiwalu Kultury Żydowskiej w Krakowie (wspólnie z dr Anną Marią Orla-Bukowską). Jest

autorem projektów ogólnopolskich badań konsumenckich oraz publikacji dotyczących wykorzystania zaawansowanych technik analizy treści w różnorodnych środowiskach CAQDAS.

Adres kontaktowy:

Instytut Socjologii
Uniwersytet Jagielloński
ul. Grodzka 52, 30-962 Kraków
e-mail: k_tomanek@wp.pl

Wprowadzenie – inspiracje teoretyczne

Badacze sięgający w praktyce po metody analizy tekstów stawiają przed nimi różnorodne cele. Od pozyskiwania prostych informacji tekstowych (*Information Extraction* [IE]¹) po odkrywanie modeli koncepcyjnych i wiedzy zawartej w tekstach (*Knowledge Discovery in Databases* [KDD]²). Od opracowania i kodowania informacji tekstowych (*Text Encoding* [TE]³) po klasyfikację (*Text Classification* [TC]⁴). W tym artykule poddam analizie dwie metody klasyfikacji tekstów. Opiszę ich właściwości oraz poddam ocenie wyniki uzyskane dzięki ich zastosowaniu. Zanim przejdę do opisu metod, usytuuję je w szerszej perspektywie metodologicznej, podając typologię metod klasyfikacyjnych stosowanych w analizach *text mining*.

Wyróżnić możemy dwa odmienne podejścia do zagadnienia klasyfikacji tekstów. Pierwsze poddaje analizie zawartości tekstów i wypowiedzi. To

¹ Zbiór technik (leksykalnych lub statystycznych, stosujących język logiki) służących do wydobywania z tekstów informacji, faktów.

² Proces wydobywania wiedzy z danych (również tekstowych) oparty za zastosowaniu różnorodnych technik analitycznych, takich jak: selekcja informacji z tekstów, pre-procesowanie danych, transformacje danych, zastosowanie technik *data mining* i *text mining*, interpretacja, ewaluacja.

³ Techniki TE to zbiór rozwiązań służących do opracowywania zawartości dokumentów. Celem zastosowania TE jest przygotowanie tekstu i struktury dokumentów tak, aby dawały one większe możliwości analityczne niż nieopracowane dokumenty tekstowe. Przykłady technik TE: tokenizacja, lematyzacja, *stemming*, filtrowanie, stop lista, indeksowanie.

⁴ Techniki TC to zbiór rozwiązań służących do strukturyzacji dokumentów, wypowiedzi lub części wypowiedzi. Metody te obejmują zarówno automatyczne, półautomatyczne, jak i manualne opracowywanie materiału tekstowego. Popularne i często stosowane techniki to: klasyfikacja oparta na indeksach, naiwny Bayes, *metoda K-najbliższych sąsiadów*, *drzewa decyzyjne*, *metody support vector machines* [SVR]. Więcej informacji o metodach klasyfikacji tekstów znaleźć można w: Hotho, Nürnberg, Paaß (2005).

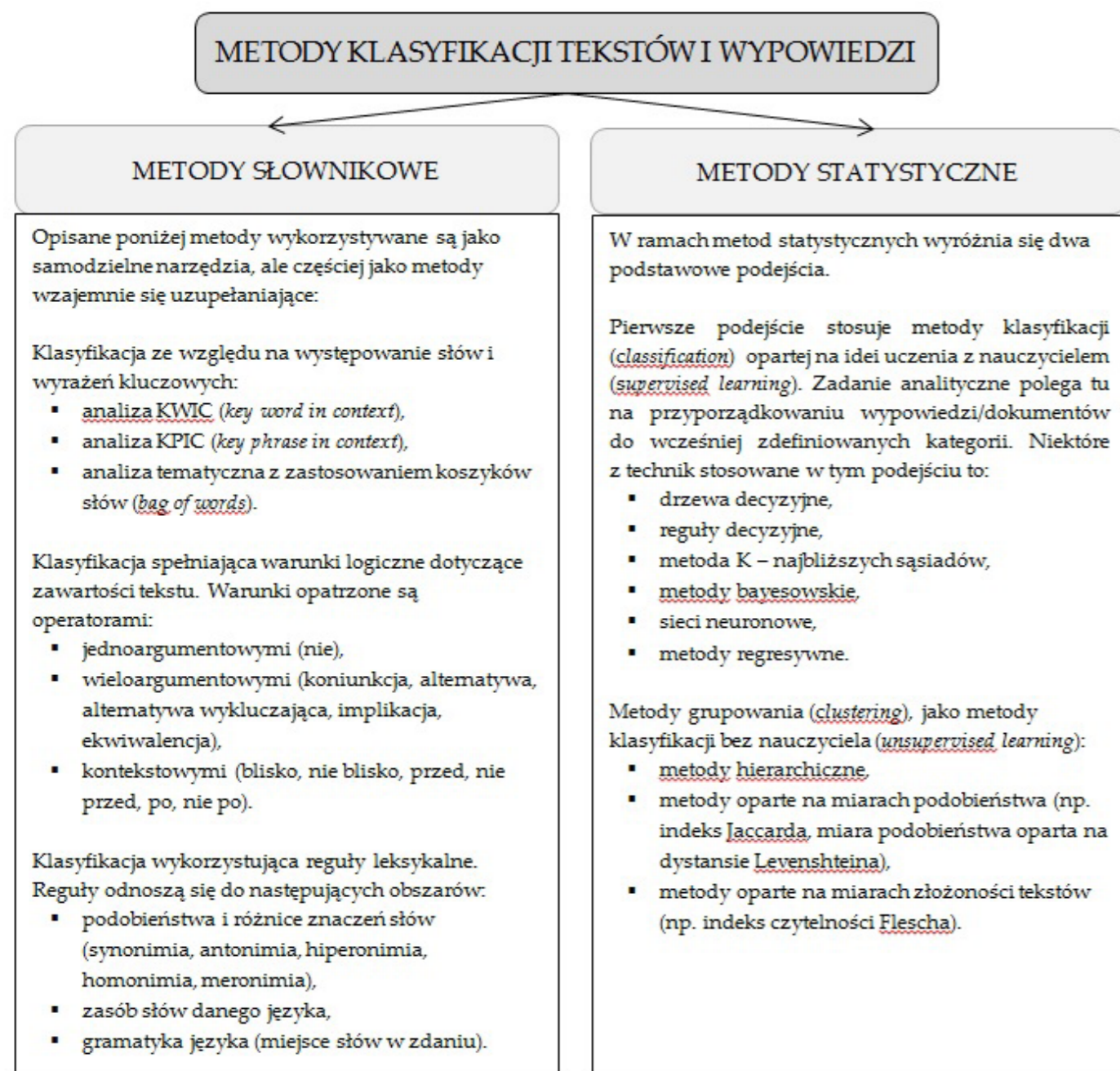
podejście wykorzystuje słowa i frazy istotne w danym tekście; posługuje się językiem logiki; wykorzystuje znaczenia analizowanych słów i frazy; bierze pod uwagę reguły leksykalne i składniowe rządzące językiem. W tym podejściu niezbędna jest znajomość: gramatyki analizowanego języka, znaczeń używanych w nim słów, specyfiki wypowiedzi związanej ze stosowanym słownictwem. Drugie podejście sięga po metody statystyczne. W tym przypadku mamy do czynienia z wykorzystaniem technik ilościowych traktujących tekst jako obiekt charakteryzowany za pomocą danych ilościowych, które opisywać mogą zarówno zawartość dokumentu (np. liczbę słów lub fraz, poziom złożoności tekstu, ilość cytowań itp.), jak i sam dokument (np. data publikacji, liczba znaków w dokumencie itp.). Tak traktowany tekst–obiekt reprezentowany jest jako wektor w wielowymiarowej przestrzeni wyznaczonej przez zbiór cech opisujących dokumenty. Poniżej podaję syntetyczny opis metod stosowanych w obrębie dwóch zarysowanych tu podejść.

W dalszej części artykułu skupię się na charakterystyce dwóch podejść. Opiszę metodę klasyfikacji słownikowej dla analizy sentymentu. Scharakteryzuję klasyfikację dokonaną za pomocą naiwnego algorytmu Bayesa.

Metoda słownikowa dla analizy sentymentu

Określenie *analiza sentymentu* odnosi się do automatycznych i półautomatycznych metod analizy tekstów. Podstawowe cele tych metod to: identyfikacja i klasyfikacja fragmentów lub całych wypowiedzi

Diagram 1. Typologia metod klasyfikacji tekstów.



Źródło: opracowanie własne.

ze względu na pojawiające się w nich słowa nacechowane emocjonalnie. Analiza sentymentu opiera się na dwóch założeniach. Po pierwsze, niektóre wypowiedziane słowa wyrażają emocje. Po drugie, istnieją słowa, których wypowiedzenie wywołać może emocje (Pang, Lee 2008). A zatem analiza sen-

tymentu z jednej strony wskazuje na stany emocjonalne autora wypowiedzi, z drugiej – służy również określeniu emocjonalnego efektu, jaki dana wypowiedź może wywołać. W tym rozumieniu termin *analiza sentymentu* wprowadzili Das i Chen (2001) oraz Tong (2001).

Analiza opinii (Pang, Lee 2008), której przykładem jest analiza sentymentu, korzysta z rozwiązań wypracowanych w obszarze przetwarzania języka naturalnego (*natural language processing* [NLP], Nasukawa, Yi [2003]). W ramach CAQDAS pierwsze próby budowy oceny sentymentu wypowiedzi pojawiły się w 2003 roku. Prekursorzy tej metody zbudowali pierwsze algorytmy dokonujące automatycznej klasyfikacji tekstów na te o pozytywnym i te o negatywnym wydźwięku emocjonalnym (Yi i in. 2003). Po tej pierwszej próbie nastąpił szybki rozwój słowników służących do analiz wypowiedzi i dokumentów (Nielsen 2011). Z jednej strony mamy do dyspozycji słowniki tematyczne⁵, klasyfikujące wypowiedzi ze względu na tematykę, której dotyczą, z drugiej strony – rozwijają się różnorodne słowniki pozwalające na identyfikację słów i wypowiedzi wyrażających lub wywołujących emocje. Słowniki te pozwalają zarówno na proste klasyfikacje (pozytywne–negatywne), ale też na klasyfikacje bardziej złożone (niepokój–chwała–agresja–smutek–miłość). Przykładem zaawansowanego słownika do analizy sentymentu może być SentiWordNet (Esuli, Sebastiani 2006). Bardziej złożony koncepcyjnie jest słownik RID (Hogenraad, Oriane 1986). Powstały również słowniki mieszane, które łączą dwie przywołane powyżej idee: analizę tematyczną oraz analizę sentymentu. Przykładem takiego narzędzia jest słownik Loughrana i McDonalda (2011), który klasyfikuje wypowiedzi odnoszące się do ekonomii, dziedziny finansów ze względu na zawarty w nich ładunek emocji.

⁵ Przykładem takiego słownika może być Visual Thesaurus klasyfikujący wypowiedzi z siedmiu obszarów tematycznych (geografia, ludzie, nauki społeczne, sztuka i literatura, matematyka, nauka, słownictwo). Por.: <http://www.visualthesaurus.com/vocabgrabber/>.

Liczba publikacji zarówno z zakresu metod, jak i zastosowań analizy sentymentu (Dini, Mazzini 2002; Cardie i in. 2003; Dave, Lawrence, Pennock 2003) jest coraz większa. Warto wspomnieć o jednej pracy – chyba najczęściej cytowanej i najlepiej znanej w tym obszarze – opublikowanej w roku 2013 analizie Acerbiego, Lamposa, Garnetta, Bentleya. Autorzy dokonali opisu literatury XX-wiecznej w duchu analizy sentymentu, a dokładniej – opisali tendencje i zmiany związane z występowaniem słów nacechowanych emocjonalnie w literaturze. Autorzy wykorzystali, co warto dodać, narzędzia z obszaru CAQDAS.

Metody analizy tekstu wspierane komputerowo służą coraz większym gronom naukowców. Powszechnie wykorzystuje się je w analizach zjawisk kultury (Michel i in. 2011), lingwistyce (Lieberman i in. 2007, historii (Pagel, Atkinson, Meade 2007), antropologii (DeWall i in. 2011).

W tym artykule opiszę zastosowanie analizy sentymentu do identyfikacji emocji w tekstach prasowych. Sprawdzę efektywność słownika RID w identyfikacji emocji i porównam ją z efektywnością metody Bayesa.

Klasyfikacja bayesowska

Probabilistyczne metody klasyfikacji tekstów oparte są na założeniu, że słowa w analizowanych tekstach zostały ze sobą zestawione w sposób losowy. W ramach tej metodologii zakłada się, że dana kategoria tekstów k_1 charakteryzuje się określonym słownictwem s_1 , a inną kategorię tekstów k_2 cechuje słownictwo s_2 . Na tej podstawie określamy

prawdopodobieństwo *a priori*, czyli decydujemy o klasyfikacji jakiegoś tekstu, nie dysponując świadectwami przemawiającymi na rzecz ani przeciw poprawności tej klasyfikacji. Zakłada się także, że tekst, który posiada słownictwo określone jako s_1 w większej liczbie niż słownictwo określone przez s_2 , powinien być zaklasyfikowany do kategorii k_1 , a nie na przykład do k_2 . Tym sposobem określamy prawdopodobieństwo *a posteriori*, czyli prawdopodobieństwo klasyfikacji w sytuacji, w której posiadamy pewne świadectwa potwierdzające poprawność klasyfikacji.

Naiwny klasyfikator Bayesa opiera się na założeniu o wzajemnej niezależności predyktorów. W przypadku klasyfikacji tekstów oznacza to, że słowa, które identyfikują określoną kategorię wypowiedzi, mogą występować niezależnie w różnych lub w tym samym tekście. Taki naiwny klasyfikator może więc identyfikować i klasyfikować słowa, nie biorąc pod uwagę kontekstu, w jakim one występują. Takie postępowanie często nie odzwierciedla specyfiki wypowiedzi tekstowej. Niemniej jednak podejście bayesowskie często okazuje się trafnym rozwiązaniem ze względu na swoją prostotę. Wzór Bayesa określa bowiem prawdopodobieństwo tego, że szanse realizacji jakiegoś zdarzenia w kolejnej próbie zależą od tego, jak często dochodzi do owego zdarzenia i jak często do niego nie dochodzi. Innymi słowy, jeśli w tekstach naukowych naiwny klasyfikator Bayesa znajdzie takie sformułowania, jak: prawdopodobieństwo, trafność przewidywań, i słowa te zdecydowanie rzadziej występowały będą w jakichkolwiek innych tekstach (np. nienaukowych), to, napotykając nowy tekst, w którym wskazane słowa wystąpią, naiwny Bayes zaklasyfikuje

go do tekstów naukowych, a nie na przykład tych z zakresu astrologii.

Najogólniej mówiąc, naiwny klasyfikator Bayesa w analizie tekstów, w wersji, w której ją prezentuję, stosuje zasadę: im więcej w analizowanym tekście zidentyfikuję słów, które zdefiniowane są w danej kategorii tekstów k_1 i jednocześnie zidentyfikuję mniej słów zdefiniowanych w kategorii k_2 , tym większe prawdopodobieństwo, że analizowany tekst należy do kategorii k_1 , z którą tekst dzieli największą liczbę słów.

W dalszej części artykułu podam przykład zastosowania algorytmu naiwnego Bayesa. Wyniki, jakie uzyskam za pomocą tej metody, porównam z efektywnością klasyfikacji opartej na metodzie słownikowej. Zanim to uczynię, podam jeszcze opis materiału, jaki zastosowałem w analizie oraz opis metod zastosowanych na etapie opracowywania tego materiału.

Źródła danych. Metody opracowywania danych

Dane wykorzystane w prezentowanej tu analizie pochodzą z grantu badawczego przyznanego przez Ministerstwo Edukacji i Szkolnictwa Wyższego⁶. Analizowany tu zbiór danych to:

- 454 artykułów prasowych opublikowanych w prasie polskiej,

⁶ Grant realizowany był w latach: 2009–2013 i poświęcony jest badaniu postaw publiczności uczestniczącej w Festiwalu Kultury Żydowskiej w Krakowie. Kierownikiem grantu jest: dr Annamaria Orla-Bukowska, grant realizują: Annamaria Orla-Bukowska, Krzysztof Tomanek.

- artykuły publikowane były w latach: 1992–2002,
- artykuły dotyczą festiwalu kultury żydowskiej o numerach: 3–6, 8–9, 11–12,
- analizą objęty jest materiał pochodzący z 61 tytułów prasowych (z czego 370 [81,5%] artykułów opublikowanych zostało w dziennikach).

Dane przekazane przez firmę monitorującą media pierwotnie składowane były w formie elektronicznej w formatach: PDF, JPEG. Pliki sprowadzone zostały do postaci tekstowej, edytowalnej. Proces ten przebiegał w następujący sposób:

- dokonana została konwersja do formatu DOCX (wykorzystano oprogramowanie Abby Fine Reader w wersji 10),
- zweryfikowano poprawność konwersji (sprawdzone zostały m.in. sposób zapisu słów, interpunkcja, kompletność informacji),
- zapis każdego tekstu jako osobnego pliku opierał się na pracy nieautomatyzowanej i był nadzorowany,
- pliki opatrzone zostały nazwami zgodnie ze zbudowaną metodą kodowania (nazwa pliku wykorzystana została jako źródło podstawowych informacji o artykule prasowym i zawierała następujące informacje: data wydania [dzień, miesiąc, rok], tytuł czasopisma, numer festiwalu kultury żydowskiej [FKŻ], którego dotyczył artykuł),

- wszystkie artykuły zostały zarchiwizowane (pliki skategoryzowane są zgodnie z numeracją FKŻ, każdy plik posiada kopię zapasową, składowaną na innym komputerze).

Dla dalszej analizy istotne są zastosowane wobec materiału prasowego techniki kodowania. Wśród zastosowanych najważniejsze to:

- kodowanie tematyczne (identyfikacja i klasyfikacja treści do zdefiniowanych uprzednio obszarów tematycznych, np. wypowiedzi dotyczące: sztuki, religii, FKŻ, relacji polsko-żydowskich itp.),
- kodowanie zogniskowane (technika polegająca na pogłębieniu analizy wykonanej za pomocą kodowania tematycznego, tak aby w kolejnej iteracji kod obejmował fragment danych tekstowych precyzyjnie odpowiadający pytaniu badawczemu, np. w ramach danych tekstowych oznaczonych kodem „sztuka” identyfikowane są fragmenty dotyczące filmu, fotografii, malarstwa, muzyki, wystaw zdjęć itp.),
- kodowanie zerojedynkowe (kodowanie nadające dokumentowi jedną z dwóch wartości, np. wypowiedź wyrażająca pozytywne emocje lub negatywne emocje – technika zwana również *dummy coding*),
- opis wielozmiennowy – specyficzne rozumienie kodowania sprowadzające się do nadania tekstom prasowym kodów identyfikujących ich przynależność do danej klasy obiektów (np. typ artykułu: wywiad, program, relacja,

zapowiedź, recenzja, relacja po zakończeniu FKŻ, inne), taki opis poddawany może być procesowi binaryzacji (por. kolejny punkt),

- kodowanie entropijne lub kontekstowe kodowanie binarne (technika binaryzacji złożonych zmiennych opisujących dane tekstowe, tak zwane *context-adaptive binary arithmetic coding* [CABAC], na przykład redukcja wektora charakteryzującego tekst za pomocą kategorii emocji w koncepcji RID do zmiennej porządkowej w postaci: wypowiedź z przewagą słów negatywnych emocjonalnie – wypowiedź neutralna – wypowiedź z przewagą słów pozytywnych emocjonalnie).

Perspektywa metodologiczna

Zastosowane przeze mnie podejście metodologiczne czerpie z kilku tradycji. Opiszę ich podstawy w sposób syntetyczny.

Najczęściej stosuję perspektywę, jaką w metodologii proponuje pragmatyzm. Charles Sanders Peirce miał nadzieję przenieść nauki laboratoryjne do filozofii, z kolei William James objawiał się jako trzeźwy empirysta, a John Dewey czerpał bezpośrednio z przyrodoznawstwa, próbując unaukować dociekania filozoficzne. Ci trzej klasycy dali podstawy do rozwiązań, jakie w filozofii i metodologii nauki proponował Richard Rorty. Metodologię naukową traktował on jak skrzynkę z narzędziami. Sięgając do niej, sięgamy po wiele rozwiązań, z których nie tylko jedno będzie przydatne w potrzebie (Rorty 1996; 1999). Stosując ideę Rortyego, staram się stosować różnorodne roz-

wiązania analityczne z obszaru analiz jakościowych i ilościowych.

Druga wykorzystywana przeze mnie strategia czerpnie z obszaru *Mixed Methods Research*. To podejście przydatne jest mi wtedy, gdy:

- weryfikuję trafność wyników analiz jakościowych (QUAL), odwołując się do metod analiz ilościowych (QUAN),
- wyjaśniam wyniki QUAN, sięgając do danych QUAL.

Trzecia tradycja, do której sięgam, to obszar metod związanych z przetwarzaniem języka naturalnego (NLP). Stąd właśnie zapożyczam ideę analizy sentymentu. W obszarze NLP idea ta jest szczególnym rodzajem analizy opinii (*opinion mining* [Pang, Lee 2008]).

Narzędzia analityczne

W analizie sentymentu posłużę się słownikiem klasyfikującym słowa w dwojaki sposób. Po pierwsze dokonuję analizy sentymentu, klasyfikując słowa do wielu kategorii. Na drugim etapie dokonuję uproszczenia i klasyfikuję słowa do dwóch kategorii: pozytywne (emocje pozytywne, uczuciowość, chwała) i negatywne (niepokój, smutek, agresja). Słownik będący podstawą klasyfikacji to RID (*Regressive Imagery Dictionary*). Autorem słownika jest profesor psychologii Colin Martindale (1976; 1977; 1990). Słownik odnosi się do podstawowych i pierwotnych procesów poznawczych, którym przypisuje specyficzne dla nich słowa. Typy kategorii, które wyróżnił Martindale, prezentuję w tabeli 1.

Tabela 1. Kategorie emocji w słowniku Colina Martindale'a.

EMOCJE	
KATEGORIE	PRZYKŁADOWE SŁOWA
Pozytywne	wzruszenie, wesołość, radość, zabawa
Niepokój	obawa, strach, fobia
Smutek	depresja, niezadowolenie, samotność
Uczuciowość	czułość, małżeństwo, miłość
Agresja	wściekłość, sarkazm, przykrość
Ekspresja	zachowania, sztuka, taniec, śpiew
Chłuba	podziw, bohaterskość, duma, król

Źródło: opracowanie własne na podstawie słownika Martindale'a dostępnego na stronie <http://provalisresearch.com>.

Stosując słownik RID, wykorzystuję narzędzia CADAS i inne narzędzia IT. Są to:

- pakiet QDA Miner/Wordstat/Simstat do zestawień i analiz statystycznych,
- darmowe środowisko IDLE Python's Integrated Development Environment do czyszczenia danych tekstowych oraz stosowania procedur deduplikacji.

Dodatkowo stosuję stop listę. Jest to lista słów, które są wykluczone z analiz. Należy wspomnieć, że wybór stop listy do analiz nie powinien nigdy odbywać się na zasadzie polecenia czy dobrej renomy listy. Każda z analiz niesie ze sobą konkretne pytania badawcze, co oznacza konieczność dostosowania do nich wykorzystywanych narzędzi badawczych, a zatem też stop listy.

Techniki i procesy analityczne

Identyfikacja i klasyfikacja słów do dwóch wyodrębnionych kategorii (pozytywne, negatywne) wykonane zostały z zastosowaniem słownika RID. Analiza ta przebiegała w pięciu etapach:

- pierwszy etap to n-gramowa klasyfikacja bez nadzorowania. N-gramowa to w tym przypadku analiza 1-gramowa, czyli polegająca na identyfikacji pojedynczych słów. Jest to najprostsze z rozwiązań metodologicznych polegające na rozpoznaniu słów w tekście oraz porównaniu ich ze słowami znajdującymi się w słowniku RID oraz, w konsekwencji, klasyfikacji słów do jednej z kategorii słownikowych. Słowa znajdujące się w słowniku to: pełne wersje słów kluczowych związanych z emocjami (np. złość identyfikuje dokładnie jedno słowo, którego

jest odwzorowaniem) oraz rdzenie słów, wobec których zastosowana została procedura *stemmingu* (proces polegający na wydobyciu z wybranego wyrazu tzw. rdzenia, a więc tej jego części, która jest odporna na odmiany przez przypadki, rodzaje itp., np. uprzejm⁷ identyfikuje i klasyfikuje takie słowa, jak: uprzejmość, uprzejmie, uprzejmy, uprzejma),

- weryfikacja sklasyfikowanych słów odbywała się poprzez analizę kontekstową i analizę nadzorowaną (nieautomatyczną). Na tym etapie poprzez analizę kontekstu, w jakim wystąpiły sklasyfikowane słowa, możliwa była ocena trafności klasyfikacji,
- trzeci etap analizy klasyfikacyjnej to rozbudowa słownika RID o reguły analityczne poddające diagnozie analizę kontekstu, w jakim występuje słowo kluczowe. Dodatkowo zakres słownika zostaje rozszerzony o analizę fraz. W ten sposób analiza n-gramowa zostaje rozszerzona o analizę fraz (pod uwagę wzięte zostały frazy dwu i trzy wyrazowe występujące więcej niż 5 razy),
- na kolejnym etapie wykonana zostaje analiza klasyfikacyjna, która bierze pod uwagę rozbudowany słownik RID i wykorzystuje bayesowski model klasyfikacji. Wyniki tej analizy zostają poddane ocenie w kolejnym kroku,

- weryfikacja użycia klasyfikatora bayesowskiego pozwala na ocenę efektywności zastosowanych metody i narzędzi analitycznych.

Wymienione tu techniki analityczne pozwoliły mi na weryfikację poprawności klasyfikacji, poprawę rzetelności analiz oraz optymalizację algorytmów klasyfikacyjnych zastosowanych w pierwotnym podejściu bez nadzorowania. Wnioski z zastosowania opisanych narzędzi i procedur omówię w ostatnim punkcie artykułu.

Wyniki analiz

Po wdrożeniu procedur czyszczenia i normalizacji danych, pierwszy etap w analizie to eksploracyjna analiza danych. W zestawieniach pojawiają się artykuły, których rok wydania nie jest znany (5 artykułów⁸). Pliki te wyłączyłem z dalszej analizy. Zastosowanie pierwszego modelu klasyfikacyjnego (bez nadzorowania) dało mi następujące wyniki:

- słowa emocjonalnie negatywne: 521,
- słowa emocjonalnie pozytywne: 400.

Uzyskany wynik poddaję weryfikacji. W pierwszej kolejności kontroluję słowa zidentyfikowane przez słownik. Weryfikuję więc trafność klasyfikacji poprzez analizę kontekstową sklasyfikowanych słów. Poniżej podaję kilka przykładów analizowanych kontekstów.

⁸ Podjąłem próby ustalenia pochodzenia wskazanych artykułów. Korespondencja z dostawcą oraz googlowanie fragmentów treści artykułów nie dały pozytywnej odpowiedzi.

Ramka 1. Przykłady analizowanych wypowiedzi.

Przykłady poprawnie sklasyfikowanych wypowiedzi pozytywnych:
 Główną postacią festiwalu był **ZNAKOMITY** trębacz Frank London (The Klezmaties, Hasidic New Wave), który poprowadził trzy koncerty.
 Jedno z największych wydarzeń już za nami - monumentalny występ **WYBITNYCH** kantorów zainaugurował tegoroczną edycję festiwalu.
 Gdy przyszło do finałowych pieśni, kantorzy zachęcili do śpiewu publiczność, która **OCHOCZO** podchwyciła nostalgiczną pieśń „Jerusalem of Gold”.

Przykłady niesklasyfikowanych wypowiedzi pozytywnych:
 Dodatkową **ATRAKCJA** wieczoru było perkusyjne trio Yakar Rhythms z Senegalu, które wystąpiło wraz z Hassidic New Wave.
AMBICJE Londona i Schwimmera, którzy towarzyszyli Sklambergowi, będącemu bez wątpienia centralną postacią koncertu, sięgały dalej. Nie komplikując nadmiernie formy, posługując się często żartem i nie niszcząc autentycznej urody i prostej **RADOŚCI** przeżywania muzyki, dźwignęli Zmirosy na wyższy, koncertowy poziom.

Źródło: opracowanie własne.

Analiza kontekstu występowania słów pozytywnych pokazała, że w paragrafach, w których widnieją słowa emocjonalnie pozytywne, istnieją też inne słowa o podobnym zabarwieniu. Ta analiza pozwala włączyć zidentyfikowane słowa do kategorii słów pozytywnych emocjonalnie. Koszyk słów pozytywnych został poszerzony między innymi o takie rdzenie słów, jak: atrakc*, ambicj*, pięk*, uroczyst*.

Klasyfikacja słów do kategorii negatywnych emocjonalnie wykazała błędne użycie niektórych ze zdefiniowanych rdzeni słów. Oto przykłady pokazujące potrzebę nowych definicji:

- eliminacja rdzenia na rzecz listy słów: groz*–groza, grozić, groźne, groźba,

- redefinicja rdzenia: cierpi*–cierpie*, zadus* –zadusi*,
- wyłączenie słów o zabarwieniu w dużej mierze zależnym od kontekstu: zabiera* (zabierać komuś czyjeś dobro, zabierać głos w dyskusji), humor.

Dodatkowe rozszerzenia słownika RID uzyskuje w efekcie dodania negacji. Tak więc nie_atrakc*, nie_lojaln* identyfikują wypowiedzi nacechowane negatywnie.

W wyniku wskazanych powyżej procedur poprawność klasyfikacji poprawiła się w sposób istotny. Różnicę pomiędzy liczbą słów zidentyfikowanych

⁷ Zapis słów z gwiazdką w roli sufiksu oznacza, że wyszukiwane są słowa, która zawierają ciąg liter umieszczony przed gwiazdką. Na przykład: dom* wyszukuje takie słowa jak: domowy, domator, domatorka, domownik. Zapis, w którym gwiazdka występuje w roli prefiksu i sufiksu, identyfikuje słowa, które zawierają ciąg litera zapisany pomiędzy gwiazdkami. Na przykład: *dom* wyszukuje takie słowa jak: domownik, zadomowiony, udomowiony, Radom.

Ramka 2. Przykłady analizowanych wypowiedzi.

Przykłady poprawnie sklasyfikowanych wypowiedzi negatywnych:

Naczytałam się w Polsce, że nie tylko Żydzi **CIERPIELI** podczas wojny, że był to kraj rozdzierany przez **ZABORCÓW**.

Tak się **STRASZNIE** zdarzyło, że naziści umieścili obozy dla Żydów na ziemi polskiej.

Inna pani, która przedstawiła się, jako Żydówka pochodząca z Krakowa, a mieszkająca w Nowym Jorku, oświadczyła, że jest jej zwyczajnie **PRZYKRO**, bowiem tamtejsi źle poinformowani Żydzi twierdzą, że cała Polska jest antysemitka.

Przykłady niepoprawnie sklasyfikowanych wypowiedzi negatywnych:

Festiwal Kultury Żydowskiej, który po raz 11 odbył się tego roku na krakowskim Kazimierzu stanowi nie lada wyzwanie dla wszelkich imprez jazzowych, jakie się dzieją w tym mieście, od pojedynczych koncertów poczynając na szacownych **ZADUSZKACH** skończywszy.

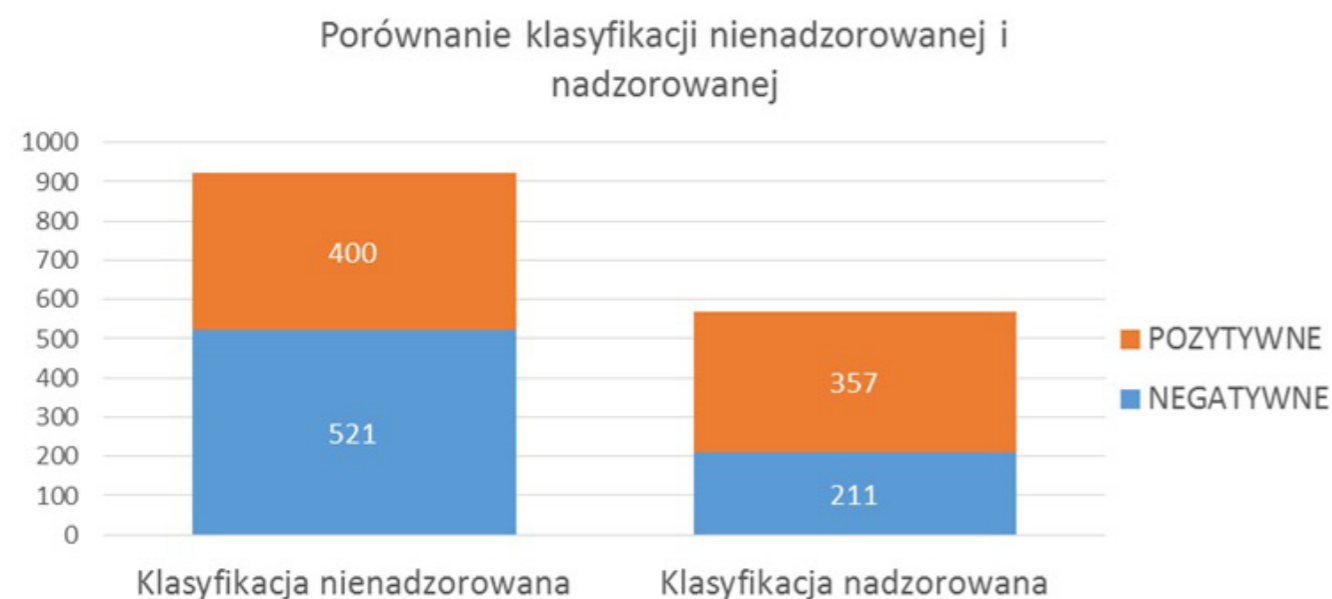
W ciągu pięciu lat namalowałem ze sto takich obrazków. Nie wystawiam ich i nie sprzedaję. Chwała Bogu, że nie muszę **ZABIEGAĆ** o pieniądze.

Wielu dyskutantów **ZABIERAŁO** głos w duchu bardzo osobistym.

Kiedyś mi zrobili piękne lakierowane trzewiki. Ubrałem je i poszedłem na futbol mecz. Kiedy futbol do mnie doszedł i ja go **KOPNAŁEM**, otworzyła mi się podeszwa od przodu do tyłu. Mama dała mi pasem. To była wspaniała kobieta.

Źródło: opracowanie własne.

Wykres 1. Porównanie rezultatów metod klasyfikacji.



Źródło: opracowanie własne.

i sklasyfikowanych za pomocą dwóch metod widać na poniższym wykresie (wykres 1).

Na tym etapie analiz dysponuję przetestowanym i zweryfikowanym podstawowym słownikiem sentymentów opartym na koncepcji Martindale'a. Skuteczność klasyfikacji z zastosowaniem tego słownika porównam teraz do trafności klasyfikacji, jaką daje klasyfikator naiwny Bayesa (Pang, Lee 2002). Jego zastosowanie do analizy tekstów zasadne jest w szczególności w sytuacji, w któ-

rej mamy podstawy, by zakładać, że pomiędzy zmiennymi opisującymi analizowane obiekty istnieje istotny związek (Domingos, Pazzani 1997). W tym przypadku jest to intuicyjne założenie mówiące, że słowa o zabarwieniu emocjonalnym występują częściej w konkretnych typach artykułów. Publikacje prasowe wykorzystane w analizie scharakteryzowane zostały w oparciu o treść, jaką zawierają. Ten zabieg doprowadził do pogrupowania tekstów w następujące kategorie (spis kategorii zawiera tabela 2):

Tabela 2. Częstość występowania zidentyfikowanych typów artykułów.

TYP ARTYKUŁU	CZĘSTOŚĆ WYSTĄPIENIA	PROCENT
zapowiedź	149	32,80%
recenzja	125	27,50%
relacja	73	16,10%
wywiad	59	13,00%
program	35	7,70%
relacja post	13	2,90%
SUMA	454	100%

Źródło: opracowanie własne.

Intuicja podpowiada, że tekst, który jest programem imprezy, zdecydowanie rzadziej zawierał będzie słowa nacechowane emocjonalnie. Dodatkowo można założyć z niewielkim błędem, że zapowiedzi mogą zawierać takich słów mniej niż recenzje i relacje post. Te wskazówki sprawiają, że klasyfikator bayesowski sprawdza się w analizie lepiej niż słow-

nik RID w jego uproszczonej wersji (uwzględniającej dwie kategorie: słowa pozytywnie i negatywnie nacechowane emocjonalnie). Podpowiedź, jaką otrzymał algorytm bayesowski, zwiększa trafność klasyfikacji paragrafów zawierających w sobie słowa o zabarwieniu emocjonalnym. Podsumowanie tej analizy podaję w tabeli 3.

Tabela 3. Porównanie analiz wykonanych za pomocą wykorzystanych metod klasyfikacji.

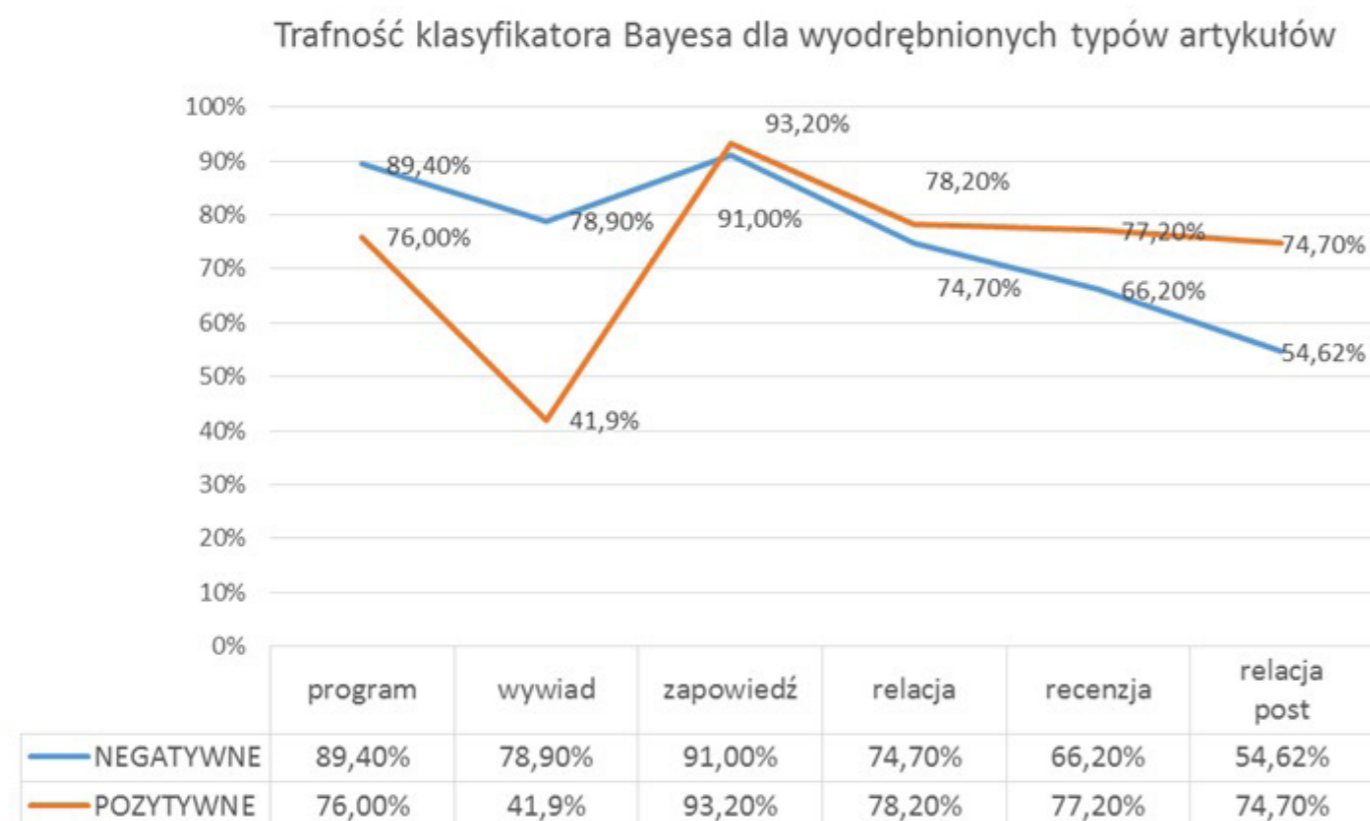
ZAKRES ANALIZY	LICZBA ZIDENTYFIKOWANYCH	NAIWNY BAYES %	NIENADZOROWANY RID %
1-gram	721	88.12	79.20
1-gram + frazy	1274	74.67	64.87

Źródło: opracowanie własne.

W szczególności większość poprawnych klasyfikacji naiwny Bayes przeprowadził w tekstach, których dotyczyło założenie o zależności pomiędzy

typem artykułu a występowaniem słów nacechowanych emocjonalnie. Wyniki klasyfikacji pokazuje wykres 2.

Wykres 2. Trafność klasyfikacji bayesowskiej w zidentyfikowanych typach artykułów (1-gram + frazy).



Źródło: opracowanie własne.

Szczególną uwagę zwraca niska trafność klasyfikacji wypowiedzi w przypadku wywiadów (pozytywne) oraz relacji post (negatywne). W tym pierwszym przypadku użycie przez rozmówców języka potocznego i spontanicznego napotyka na wiele niesklasyfikowanych w słownikach słów oraz problemy interpretacyjne, co w konsekwencji powoduje błędne klasyfikacje. Relacje post (po

zakończeniu festiwalu kultury żydowskiej) bogate były w sformułowania wieloznaczne. Dodatkowo zawarte w nich refleksje sprzyjały pojawianiu się obok siebie ocen biegunowych w jednym zdaniu oraz takich, które formułowane były z wielu punktów widzenia (np. radość publiczności i zła organizacja). Przykłady takich wypowiedzi podaje poniżej:

Ramka 3. Przykłady analizowanych wypowiedzi.

Wypowiedzi niejednoznaczne i problematyczne w klasyfikacji

WYWIAD:
 Co prawda nie dzwoniły telefony, co chwilę, ale zdarzył się jeden czy drugi cham czy chamka, który oczywiście jej nie wyłączył. Także **WYBITNIE** zakłócały odbiór tego koncertu raz po raz, często w najmniej oczekiwanych i potrzebnych momentach, rozbłyskujące flesze.

Na koncercie **KLASKAŁAM**, ale podczas modlitwy - **PŁAKAŁAM**.

Odkrywam nie tylko kulturę **CIERPIENIA**- poznaję także kulturę **OSIĄGNIĘĆ**.

RELACJE POST:
PRZECIWNIK kary śmierci.
KOCHANI wróćmy do normalności - postuluję, przynajmniej na niektórych koncertach, przynajmniej takich jak ten - zostawcie aparaty fotograficzne w domu
 Taki to był Kazimierz. **ZACNY** i podły. Smutny i wesoły. Na jednym jego końcu tłum zęgnął zmarłego rabina **Komitzcra**, a na drugim stary chasyd spieszył się na ślub córki cadyka z Bobowej.

Źródło: opracowanie własne.

Wskazane powyżej, zidentyfikowane w trakcie analiz, błędy klasyfikacyjne oraz modyfikacje słowników wykorzystałem w celu przebudowania słownika RID. Aby pokazać i zaakcentować wagę tego zabiegu, podam klasyfikację wypowiedzi prasowych dokonaną za pomocą słownika RID w jego pierwotnym kształcie oraz klasyfikację dokonaną po jego modyfikacjach.

W pierwotnej postaci słownik RID stanowi strukturę siedmiu kategorii⁹. Klasyfikacja tekstów prasowych za pomocą niezmienionego RID pokazuje, że analizowane teksty zawierają w przeważającej ilości wyrażenia i słowa związane

⁹ Nazwy kategorii zmodyfikowałem tak, aby lepiej oddawały ich zawartość. Akceptowalnym kosztem tej modyfikacji jest wydłużenie nazw kategorii.

z zachowaniami ekspresyjnymi oraz te odwołujące się do chwały i chluby. Te dwie kategorie określają 69,9% zidentyfikowanych wypowiedzi prasowych.

Tabela 4. Klasyfikacja tekstów prasowych przez niezweryfikowany słownik RID.

	CZĘSTOŚĆ WYSTĘPOWANIA	%	LICZBA TEKSTÓW	%	TRAFNOŚĆ KLASYFIKACJI %
EKSPRESJA	1779	39,6	376	82,3	64,5
CHLUBA/CHWAŁA	1317	29,3	356	77,9	45,2
UCZUCIOWOŚĆ	474	10,6	219	47,9	41,1
POZYTYWE ODCZUCIA	310	6,9	172	37,6	29,8
AGRESJA/PRZEMOC/NIECHĘĆ	233	5,2	132	28,9	62,2
NIEPOKÓJ	207	4,6	134	29,3	61,8
SMUTEK	128	2,9	82	17,9	54,2

Źródło: opracowanie własne.

Modyfikacja słownika RID w zakresie trafności klasyfikacji, poprawności zapisów rdzeni słów, trafności skategoryzowanych fraz oraz analizy kontekstu występowania zidentyfikowanych słów pozwala na osiągnięcie nowego wyniku. Po pierwsze, zmodyfikowany słownik identyfikuje mniejszą liczbę wypowiedzi. Po drugie, zwiększa się istotnie procent poprawności klasyfikacji. Dane te pokazują w tabeli 5.

Tabela 5. Klasyfikacja tekstów prasowych po weryfikacji słownika RID.

	CZĘSTOŚĆ WYSTĘPOWANIA	%	LICZBA TEKSTÓW	%	TRAFNOŚĆ KLASYFIKACJI %
EKSPRESJA	987	0,20	365	27%	73,8
CHLUBA/CHWAŁA	911	0,18	328	24%	77,2
UCZUCIOWOŚĆ	427	0,09	211	15%	69,5
POZYTYWE ODCZUCIA	301	0,06	153	11%	54,3
NIEPOKÓJ	198	0,04	121	9%	78,8
AGRESJA/PRZEMOC/NIECHĘĆ	193	0,04	124	9%	69,9
SMUTEK	72	0,01	71	5%	72,1

Źródło: opracowanie własne.

Wnioski i rekomendacje

Zagadnienia związane z automatyczną analizą treści wykraczają szeroko poza metody dyskutowane w tym artykule¹⁰. W tym obszarze problemowym mierzę się zaledwie z jednym zagadnieniem. Jest nim porównanie trafności klasyfikacji tekstów prasowych za pomocą dwóch metod (metoda słownikowa z zastosowaniem RID, naiwny klasyfikator Bayesa). W trakcie analiz obie metody ujawniły swoje słabe i silne strony, które teraz opiszę.

RID wykazuje względną skuteczność klasyfikacyjną, kiedy wzbogacony zostaje o automatyczne metody klasyfikacji. Naiwny Bayes wydaje się być dobrym punktem wyjścia dla analizy sentymentu. To podejście wymaga dodatkowego etapu, którym jest uczenie nadzorowane.

Na przykładzie przeprowadzonej analizy można sformułować hipotezę brzmiącą: RID niewzbogacony o reguły analiz kontekstowych wykazuje względnie słabą trafność klasyfikacyjną w języku polskim w przypadku wypowiedzi spontanicznych i w analizie języka niesformalizowanego (np. wywiady). A zatem efektywność klasyfikacji za pomocą RID może być zależna od typu tekstu i typu języka, wobec których słownik ten jest stosowany.

Dwa przywołane powyżej wnioski wzmacniają twierdzenie o potrzebie ewaluacyjnego podejścia do klasyfikacyjnych analiz tekstów. Przed dokonaniem analizy, procedurom ewaluacji poddane powinny być takie narzędzia, jak: słownik klasyfikacyjny,

¹⁰ Szersze omówienie tych zagadnień znaleźć można w Jurski, Martin (2009).

reguły leksykalne słownika, stop lista, algorytm lematyzacji, rdzenie słów. Dodatkowo proces budowy i doskonalenia słowników klasyfikacyjnych powinien być poddany procedurze wielokrotnej weryfikacji reguł słownikowych. Ten zabieg stosowany w trakcie analiz pozwala na znaczące zwiększenie poprawności klasyfikacji.

W automatycznych analizach tekstów niezbędną staje się miara nietrafności klasyfikacji. Zagadnienie to dotyczy niepewności i błędu pomiaru. Miara błędnej klasyfikacji jest domeną skwantyfikowanych analiz tekstowych (Hopkins, King 2010). Niemniej jednak ocena błędu klasyfikacji może być stosowana również przez metody nieprobabilistyczne, a wśród nich przez metody słownikowe.

Wynik klasyfikacji osiągnięty za pomocą metody Bayesa może być zweryfikowany przez reguły słownikowe. Te dwie metody w zdecydowanie krótszym czasie mogą dać poprawniejszy wynik klasyfikacji niż każda z nich stosowana z osobna.

Poza testowanym słownikiem RID istnieje kilka innych słowników do analizy sentymentu. Wartym sprawdzenia jest podejście, które testowałoby trafność zastosowania różnych słowników do analizy sentymentu dla danego rodzaju wypowiedzi tekstowych (artykułów prasy codziennej, artykułów prasy branżowej, języka subkultury itp.).

Na koniec chciałbym dodać bardziej ogólną refleksję dotyczącą analizy sentymentu. W zdecydowanej większości przypadków opiera się ona na identyfikacji słów i fraz kluczowych. Podejście takie bez wiedzy o zwyczajach językowych autorów

wypowiedzi, uwzględnienia specyfiki użycia słów, kontekstowej zmienności znaczeń, ale też bez informacji o sposobie wypowiedzi (np. tonie głosu autora wypowiedzi, czym zajmuje się obszar zwany *voice mining*) nastęrcza wiele trudnych do rozwiązania problemów. W szczególności pojawiają się one w analizie wypowiedzi spontanicznych oraz wypowiedzi formułowanych w języku niesformalizowanym. W najlepszym wypadku opisane w artykule metody dają poprawność klasyfikacji wypowiedzi

Bibliografia

Acerbi Alberto i in. (2013) *The Expression of Emotions in 20th Century Books*. „PLoS ONE”, vol. 8, no. 3, s. 1–6 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <<http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0059030&representation=PDF>>.

Cardie Claire i in. (2003) *Combining low-level and summary representations of opinions for multi-perspective question answering* [w:] *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*, s. 20–27 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <<http://www.aaai.org/Papers/Symposia/Spring/2003/SS-03-07/SS03-07-004.pdf>>.

Das Sanjiv R., Chen Mike J. (2001) *Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web*, „Management Science”, Vol. 53, No. 9, s. 1375–1388 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <http://algo.scu.edu/~sanjivdas/chat_FINAL.pdf>.

Dave Kushal, Lawrence Steve, Pennock David M. (2003) *Mining the peanut gallery: Opinion extraction and semantic classifi-*

na poziomie 80%. Trzeba jednak dodać, że prezentowane tu podejście biorące pod uwagę frekwencje występowania słów nie jest ani jedynym, ani najbardziej trafnym z istniejących. Czym innym jest bowiem zliczenie słów wyrażających emocje, a czym innym rozumienie wypowiedzi przez pryzmat intencji autora wypowiedzi. Interesującym byłoby zatem zaprojektowanie i wykonanie analizy, która pozwoliłaby na porównanie wyników uzyskanych za pomocą dwóch wspomnianych tu podejść.

cation of product reviews [w:] *Proceedings of WWW*, s. 519–528, [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <<http://www.kushaldave.com/p451-dave.pdf>>.

DeWall Nathan C. i in. (2011) *Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics*. „Psychology of Aesthetics, Creativity, and the Arts”, vol. 5, no. 3, s. 200–207.

Dini Luca, Mazzini Giampaolo (2002) *Opinion classification through information extraction* [w:] *Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining)*, s. 299–310 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <http://www.google.pl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fia2010primercuat.googlecode.com%2Fsvn-history%2F45%2Ftrunk%2FSEI-GO%2Fdocs%2F10.1.1.109.1736.pdf&ei=D6diU9ahG8ep7AbGu4GYDQ&usg=AFQjCNGlZrQDMZ3aj-M_a-Yv4ITbwdU0KQ&bvm=bv.65788261,d.ZGU&cad=rja>.

Domingos Pedro, Pazzani Michael (1997) *On the optimality of the simple Bayesian classifier under zero-one loss*. *Machine Learning*, vol. 29, no. 2/3, s. 103–130.

Esuli Andrea, Sebastiani Fabrizio (2006) *SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining* [w:] *Proceedings of the 5th Conference on Language Resources and Evaluation, LREC'06*, s. 417–422 [dostęp 1 maja 2014]. Dostępny w Internecie: <<http://gandalf.aksis.uib.no/lrec2006/pdf/384.pdf>>.

Hogenraad Robert, Orianne Emilie (1986) *Imagery, regressive thinking, and verbal performance in internal monologue*. „Imagination, Cognition, and Personality”, vol. 5, no. 2, s. 127–145.

Hopkins Daniel, King Gary (2010) *Extracting systematic social science meaning from text*. „American Journal of Political Science”, vol. 54, no. 1, s. 229–247.

Hotho Andreas, Nürnberger Andreas, Paaß Gerhard (2005) *ABrief Survey of Text Mining*. „German Journal for Computer Linguistics and Speech Technology”, vol. 20, no. 1, s. 19–62.

Jurafsky Dan, Martin James H. (2009) *Speech and natural language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Lieberman Erez i in. (2007) *Quantifying the evolutionary dynamics of language*. „Nature”, vol. 449, no. 7163, s. 713–716.

Loughran Tim, McDonald Bill (2011) *When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks*. „The Journal of Finance”, vol. 66, no. 1, s. 35–65.

Martindale Colin (1976) *Primitive mentality and the relationship between art and society*. „Scientific Aesthetics”, vol. 1, s. 5–18.

----- (1977) *Syntactic and semantic correlates of verbal tics in Gilles de la Tourette's syndrome: A quantitative case study*. „Brain and Language”, vol. 4, s. 231–247.

----- (1990) *The clockwork muse: The predictability of artistic change*. New York: Basic Books.

Michel Jean-Baptiste in. (2011) *Quantitative Analysis of Culture Using Millions of Digitized Books*. „Science”, vol. 331, s. 176–182.

Nasukawa Tetsuya, Yi Jeonghee (2003) *Sentiment analysis: Capturing favorability using natural language processing* [w:] *Proceedings of the Conference on Knowledge Capture (K-CAP)* s. 70–77 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <http://tredocs.com/tw_files2/urls_41/40/d-39217/7z-docs/7.pdf>.

Nielsen Finn Å. (2011) *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs* [w:] RoweMatthew i in., eds., *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings*, Heraklion, s. 93–98 [dostęp 1 maja 2014 r.]. Dostępny w Internecie: <http://ceur-ws.org/Vol-718/msm2011_proceedings.pdf>.

Pagel Mark, Atkinson Quentin D., Meade Andrew (2007) *Frequency of word-use predicts rates of lexical evolution throughout Indo-European history*. „Nature”, vol. 449, s. 717–720.

Pang Bo, Lee Lillian (2002) *Thumbs up? Sentiment Classification using Machine Learning Techniques*. „EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing”, vol. 10, s. 79–86.

----- (2008) *Opinion Mining and Sentiment Analysis*. „Foundations and Trends in Information Retrieval”, vol. 2, s. 1–135.

Rorty Richard (1996) *Przygodność, ironia i solidarność*. Przełożył Waclaw J. Popowski. Warszawa: Spacja.

----- (1999) *Obiektywność, relatywizm i prawda*. Przełożył Janusz Margański. Warszawa: Aletheia.

Tong Richard M. (2001) *An operational system for detecting and tracking opinions in on-line discussion* [w:] *Working Notes of the SIGIR Workshop on Operational Text Classification*. New York: ACM, s. 1–6.

Yi Jeonghee i in. (2003) *Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques* [w:] *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*. Washington: IEEE Computer Society, s. 427–434.

Cytowanie

Tomanek Krzysztof (2014) „Analiza sentymentu” – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 118–136 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

Sentiment Analysis. An Example of Application and Evaluation of RID Dictionary and Bayesian Classification Methods in Qualitative Data Analysis Approach

Abstract: The purpose of this article is to present the basic methods for classifying text data. These methods make use of achievements earned in areas such as: natural language processing, the analysis of unstructured data. I introduce and compare two analytical techniques applied to text data. The first analysis makes use of thematic vocabulary tool (sentiment analysis). The second technique uses the idea of Bayesian classification and applies, so-called, naive Bayes algorithm. My comparison goes towards grading the efficiency of use of these two analytical techniques. I emphasize solutions that are to be used to build dictionary accurate for the task of text classification. Then, I compare supervised classification to automated unsupervised analysis' effectiveness. These results reinforce the conclusion that a dictionary which has received good evaluation as a tool for classification should be subjected to review and modification procedures if is to be applied to new empirical material. Adaptation procedures used for analytical dictionary become, in my proposed approach, the basic step in the methodology of textual data analysis.

Keywords: qualitative data analysis, sentiment analysis, content analysis, text mining, coding techniques, natural language processing, RID dictionary, naive Bayes, CAQDAS

Jan Winczorek

Uniwersytet Warszawski

DOI: <https://doi.org/10.18778/1733-8069.10.2.08>**Wykorzystanie oprogramowania R i RQDA w jakościowo-ilościowej analizie treści orzeczeń Trybunału Konstytucyjnego**

Abstrakt Artykuł opisuje wykorzystanie pakietu statystycznego „R” z nakładką CAQDA „RQDA” w jakościowo-ilościowej triangulowanej analizie treści z kodowaniem swobodnym, analizą rzetelności kodowania i statystycznej istotności wyników, przeprowadzonej na próbie uzasadnień orzeczeń Trybunału Konstytucyjnego z lat 1986–2009. Przedstawia najważniejsze własności oprogramowania, założenia badania, jego metodologię i zastosowane procedury analizy danych, a także pytania epistemologiczne, które pojawiły się w efekcie jego realizacji.

Słowa kluczowe CAQDA, RQDA, R (oprogramowanie), analiza treści, rzetelność kodowania, Trybunał Konstytucyjny

Niniejszy artykuł przedstawia założenia metodologiczne, doświadczenia realizacyjne i niektóre wnioski badawcze powzięte w efekcie realizacji badania „Wzorce wykładni konstytucji w orzecznictwie Trybunału Konstytucyjnego w la-

tach 1986–2009”¹. Badanie to zostało zrealizowane w latach 2010–2012 przez sześcioposobowy zespół badawczy pracowników Wydziału Prawa i Administracji Uniwersytetu Warszawskiego² i miało charakter interdyscyplinarny. Wychodziło od ogólnych

Jan Winczorek, dr, adiunkt w Katedrze Socjologii Prawa Wydziału Prawa i Administracji Uniwersytetu Warszawskiego. Zajmuje się socjologią prawa, socjologią teoretyczną i teorią prawa. Opublikował między innymi książkę poświęconą prawu w teorii Luhmanna i kilkadziesiąt innych tekstów.

Adres kontaktowy:

Wydział Prawa i Administracji
Uniwersytet Warszawski
ul. Krakowskie Przedmieście 26/28
00-926 Warszawa
e-mail: janwin@janwin.info

¹ Realizacja badania opisywanego w artykule została dofinansowana z grantu NCN N N110 121237, którego realizacją kierował dr hab. Tomasz Stawecki, prof. UW. Przedstawienie wyników zawierać będzie praca zbiorowa pod redakcją prof. Tomasza Staweckiego imożą, która zostanie opublikowana w 2014 roku w wydawnictwie Wolters Kluwer. Niniejszy tekst koncentruje się wyłącznie na wątkach metodologicznych; ze względu na znaczną objętość nie jest możliwe przedstawienie w nim nawet cząstkowych wyników badania.

² W skład zespołu wchodził: dr hab. Tomasz Stawecki, prof. UW, dr hab. Marcin Matczak, dr Wiesław Staśkiwicz, dr Krzysztof Kaleta, mgr Marcin Romanowicz, dr Jan Winczorek. Podział prac w ramach zespołu przedstawiał się następująco: J. Winczorek – założenia metodologiczne badania w głównej części, opracowanie narzędzi informatycznych, wykonanie prac badawczych na części materiału w objętości wynikającej z przyjętej metodologii badania, analiza statystyczna danych, raport z wyników; pozostali członkowie zespołu – współautorstwo (w mniejszej części) założeń metodologicznych, wykonanie prac badawczych na części materiału w objętości wynikającej z przyjętej metodologii badania.

obserwacji dokonywanych w teorii prawa i teorii polityki na temat pozycji sądownictwa konstytucyjnego w systemie prawnym i ustroju politycznym współczesnego społeczeństwa. Głównym obiektem zainteresowania były w nim metody dyskursywnego uzasadniania decyzji wykorzystywane przez polski Trybunał Konstytucyjny. Przedmiot badania stanowiły więc stwierdzenia wypowiedziane w języku ściśle profesjonalnym, artykułowane za pomocą kategorii i schematów rozumowania słabo znanych poza światem prawniczym. Realizacji tego celu poznawczego służyła natomiast technika badawcza wykorzystywana częściej w naukach społecznych niż w naukach prawnych – jakościowo-ilościowa, wspomaganą komputerowo analiza treści.

Artykuł jest podzielony na cztery części. Rozpoczynają go uwagi wstępne, w których przedstawione zostają założenia badania i najważniejsze inspiracje jego autorów. W części drugiej omawiane są główne cechy i funkcje oprogramowania (programu statystycznego wraz ze specjalizowaną nakładką *Computer-Assisted Qualitative Data Analysis*, dalej CAQDA) użytego do realizacji badania. W części trzeciej mowa jest o zastosowanych w badaniach procedurach analizy danych. W części czwartej przedstawiane są uwagi na temat granic produktywności narzędzi CAQDA w socjologicznych analizach orzecznictwa sądowego i ogólniejszych problemów epistemologicznych, jakie się z tym wiążą.

Uwagi wstępne

Zagadnienie metod wykładni prawa służących do uzasadniania decyzji sądowych stanowi, wraz z pokrewnymi kwestiami tak zwanej ideologii stosowa-

nia prawa oraz argumentacji i rozumowań prawniczych, jeden z centralnych tematów teorii prawa. W ramach tej dyscypliny wypracowano wiele normatywno-proskryptywnych doktryn wykładni uzasadniających rozmaite praktyki podejmowane w tym obszarze (przydatny przegląd stanowisk zawiera książka Feteris [1999]; w sprawie bardziej idiosynkratycznych koncepcji formułowanych w polskim kontekście zob. np. Wróblewski 1988; Zieliński 2008; Morawski 2010). Rzadziej spotyka się – choć i one są tworzone w naukach prawnych – deskryptywne teorie odnoszące się do rzeczywistego przebiegu procesów decyzyjnych, zakresu władzy sędziowskiej i granic jej dyskrecjonalności. Wykładnia i stosowanie prawa stanowi również obiekt zainteresowania nauk społecznych – na przykład w kontekście zjawiska „judycjalizacji polityki” dostrzeganego przez nauki polityczne (Stone Sweet 1999; Maravall 2003) czy złożoności czynników wpływających na ostateczną treść wyroku, obserwowanych przez socjologię czy antropologię prawa (za przykłady badań prowadzonych metodami jakościowymi mogą posłużyć prace Lautmanna [1972] czy Scheffera [2010]).

Sądy konstytucyjne, czy szerzej – instytucje orzekające o tak zwanej hierarchicznej zgodności aktów prawnych, są tu obiektem wzmożonego zainteresowania (szerokie omówienie literatury w kontekście polskim i zagranicznym można znaleźć w pracy pod red. Staweckiego i Winczorka [w druku]). Wynika ono, z jednej strony, ze szczególnego charakteru i szerokiego zakresu władzy, jaką dysponują te organizacje, oraz, z drugiej strony, z ich specyficznego instytucjonalnego usytuowania *vis à vis* innych organów władzy. Pierwsza z tych kwestii

wiąże się z – z natury rzeczy – szerokimi formalnoprawnymi uprawnieniami tych ciał i ich znaczną dyskrejoną władzą orzeczniczą, wynikającą z abstrakcyjnej i niekiedy niejasnej materii konstytucyjnej, na podstawie której są zmuszone wydawać swoje wyroki. Nie tylko więc sądy konstytucyjne są jedynymi, obok parlamentów, organami uprawnionymi do uchylania norm powszechnie obowiązującego prawa, ale także wykonują swoją władzę na podstawie bardzo nieostrych przesłanek. W związku z tym dysponują zdolnością określaną w nauce prawa mianem *kompetenz-kompetenz*, to znaczy faktyczną możliwością określania własnych uprawnień poprzez odpowiednią interpretację przepisów, na podstawie których orzekają. Dostrzegając ten stan rzeczy, wysuwa się często pod ich adresem apele o stosowanie się do ideału „powściągliwości sędziowskiej”.

Gdy chodzi o drugie wspomniane zagadnienie, w obfitej literaturze przedmiotu wskazuje się na trudności, jakie teoria demokracji ma z uzasadnieniem istnienia trybunałów konstytucyjnych (Sadurski 2008: 21 i nast.). W tak zwanym skoncentrowanym modelu kontroli konstytucyjności, funkcjonującym dziś także w Polsce, trybunały konstytucyjne są ciałami quasi-sądowymi, niemieszczącymi się w klasycznym trójpodziale władzy. Orzekający w nich sędziowie pochodzą z nominacji politycznych, tak samo jak osoby wchodzące w skład władzy ustawodawczej i wykonawczej. Jednocześnie jednak zakres ich rozliczalności jest bardzo ograniczony, podobnie jak ma to miejsce w przypadku sędziów sądów powszechnych – które są zazwyczaj ciałami merytokratyczno-technokratycznymi. W związku z tym nie jest rzeczą jasną, na jakiej

podstawie możliwe jest legitymizowanie decyzji podejmowanych przez trybunały.

W klasycznej dla Europy tak zwanej austriackiej doktrynie sądownictwa konstytucyjnego problem legitymizacji rozwiązuje się, podkreślając szczególność funkcji pełnionej przez trybunały konstytucyjne (zob. Wronkowska 2008; Kelsen 2009). Mają one być wyłącznie „negatywnym ustawodawcą”, to znaczy ciałem uprawnionym jedynie do uchylania wadliwych (niekonstytucyjnych) norm prawa, ale nie tworzenia nowych. W rzeczywistości władza interpretowania prawa, z której korzystają trybunały, a także istotne problemy praktyczne (jak na przykład zagadnienie tzw. zaniechania ustawodawczego) sprawiają, że organy te nie uchylają po prostu norm prawnych, ale pełnią także rolę „pozytywnych ustawodawców”, to znaczy wydając swoje orzeczenia, *de facto* tworzą nowe normy.

Zarówno jedna, jak i druga kwestia pozostaje w bliskich związkach z zagadnieniem sposobu uzasadniania decyzji sądowych podjętym w omawianych badaniach. Można mianowicie twierdzić, że problem legitymizacji daje się przynajmniej częściowo rozwiązać wtedy, gdy działalność trybunałów będzie miała charakter dyskursywny, to znaczy nie będzie opierać się tylko na czysto formalnych uprawnieniach do podejmowania określonych decyzji, ale będzie również zmierzała do przekonania ich audytorium o trafności wydawanych wyroków (jedną z bardziej znanych doktryn tego rodzaju przedstawił Alexy [1985; 1996]). Rozważając zagadnienie dyskrejonowej władzy trybunałów, zauważa się natomiast, że realizacja wymogu intersubiektywnej uzasadnialności wyroków w istotny sposób

ogranicza arbitralność podejmowanych decyzji, ponieważ nie każda decyzja może zostać przekonująco uzasadniona w danym kontekście kulturowym, aksjologicznym czy instytucjonalnym.

Badania opisywane w niniejszym artykule zostały pomyślane tak, aby udzielały odpowiedzi na pytanie, jak w rzeczywistości swoje decyzje uzasadniał polski Trybunał Konstytucyjny (dalej TK). Nie miały jednak ambicji konkluzywnego rozstrzygnięcia problemów teorii prawa. Zmierzały natomiast do ustalenia rzeczywistych wzorców argumentowania w uzasadnieniach orzeczeń wydawanych przez TK³ oraz zbadania ich zmienności w czasie i w zależności od dostępnych zmiennych wyjaśniających. Choć więc nie były wolne od założeń teoretycznych, to miały charakter zdecydowanie eksploracyjny, nie zaś eksplanacyjny i nie miały służyć do testowania hipotez.

Sprawia to, że zanim przejdzie się do przedstawienia bliższego opisu procedury badawczej i oprogramowania użytego do jej wdrożenia, trzeba poczynić dodatkowe uwagi wstępne. Jakkolwiek orzecznictwo sądowe stanowi stały obiekt badań nauk prawnych, to już jego badanie metodami nauk społecznych nie jest powszechne. Jest tak z przynajmniej dwóch powodów. Po pierwsze, tak zwane dogmatyczne nauki prawne (na przykład nauka prawa cywilnego czy prawa administracyjnego) zorientowane są przede wszystkim na analizę zawartości

³ Warto podkreślić, że czym innym jest sposób uzasadnienia wyroku sądowego przedstawiony w jego uzasadnieniu, a czym innym rzeczywisty przebieg procesów psychicznych i społecznych, które przyczyniły się do wydania wyroku w określonym kształcie. Omawiane badanie miało służyć, rzecz jasna, ustaleniu tej pierwszej okoliczności.

źródeł prawa w celu ustalenia jego właściwej treści w świetle zewnętrznie narzuconych założeń ontologicznych i aksjologicznych. Rekonstrukcje „linii orzeczniczych” albo odwołania do pojedynczych orzeczeń mają, w polskiej kulturze prawnej i w odniesieniu do polskich sądów, charakter zazwyczaj pomocniczy. Analiza orzecznictwa służy tu głównie porównaniu rozumowania stojącego za stosowaniem prawa z tym, za którym opowiada się badacz, albo ma wprost charakter krytyczny. Z tego powodu badania zmierzające do ustalenia charakterystyki wypowiedzi sądów w dłuższym przekroju czasowym przeprowadzane na systematycznych (zwłaszcza losowych) próbach i w sposób możliwy do uogólnienia na cały korpus orzeczeń danego typu, choć mogą być istotne dla teorii prawa, nie odpowiadają potrzebom poznawczym dogmatycznych nauk prawnych.

Po drugie, prowadzenie tego rodzaju badań jest trudniejsze niż analiza wielu materiałów innego rodzaju. Wypowiedzi sądów zmierzające do uzasadnienia decyzji wyrażonych w sentencjach wyroków cechuje wysoki poziom skomplikowania, niekiedy znaczna objętość⁴ oraz pewna hermetyczność. Wynika to tyle ze stosowanej prawniczej terminologii i niekiedy niebagatelnej złożoności rozważanych zagadnień, co z usytuowania sądów w ustroju państwa i swoistej ekonomii orzekania, wpływającej na sposób przedstawiania swoich racji przez organy orzecznicze. Stwierdzenia wypowiedziane w uzasadnieniach orzeczeń mają wprawdzie – z samej zasady – charakter argumentacyjny,

⁴ W polskiej kulturze prawnej pisemne uzasadnienia orzeczeń bywają bardzo obszerne, a w przypadku Trybunału Konstytucyjnego liczą niekiedy nawet kilkadziesiąt stron.

bo zmierzają do uzasadnienia decyzji, którą podjął sąd, jednak repertuar wykorzystywanych w nich środków retorycznych jest ograniczony. Wpływają na to akceptowane *explicite* i *implicite* ideologie i metodologie orzekania oraz spodziewane reakcje stron postępowania i zapatrywania sądów wyższych instancji wyrażane podczas ewentualnej kontroli instancyjnej.

Analiza tego typu materiałów wymaga w związku z tym szczególnych kompetencji i odpowiednich narzędzi badawczych. Badacze podejmujący analizę orzecznictwa sądowego powinni w szczególności być w stanie w sposób kompetentny dekodować wypowiedzi sądów dokonywane w języku prawniczym, co wymaga znajomości doktryny prawa w konkretnym obszarze oraz powszechnie akceptowanych w danej kulturze prawnej metod argumentacji czy wnioskowań prawniczych. Jednocześnie, badacze powinni być w stanie czynić to w sposób refleksyjny, dostrzegając uwikłanie wypowiedzi sądów w społeczny kontekst. Ten drugi aspekt istotnie odróżnia socjologiczne badanie orzecznictwa od prawnego-dogmatycznego namysłu nad jego treścią i powinien być uwzględniony przy konstruowaniu narzędzi badawczych. Jest to zarazem istotną barierą dla podejmowania tego typu badań przez prawników. Konieczność posiadania kompetencji prawniczych sprawia natomiast, że badania takie są trudne do realizacji przez socjologów i innych badaczy z nauk społecznych⁵.

⁵ Nie jest przypadkiem, że w badaniach sądów i orzecznictwa najczęściej podejmowanych w naukach społecznych poszukuje się mierzalnych wskaźników, za pomocą których można w sposób zewnętrzny opisywać działania sądów (np. częstości apelacji, liczby zdań odrębnych itp.). Bardzo rzadko, jeśli kiedykolwiek, wskaźniki takie dotyczą sposobu uzasadniania orzeczeń.

Można sądzić, że opisane względy sprawiają, że badania przedstawiane w niniejszym artykule mają charakter względnie nowatorski. W Polsce podobne przedsięwzięcia badawcze (w odróżnieniu od tak zwanych badań aktowych, skupiających się zazwyczaj na treściach decyzji sądowych, a nie ich uzasadnieniach) podejmowano tylko sporadycznie. Pod koniec poprzedniej dekady trzej autorzy wchodzący w skład zespołu realizującego omawiane tu badania przeprowadzili zbliżone studia nad orzecznictwem Trybunału Konstytucyjnego, Naczelnego Sądu Administracyjnego i Sądu Najwyższego (Winczorek, Stawecki, Staśkiewicz 2008). Inny członek zespołu badawczego prowadził badania orzecznictwa Naczelnego Sądu Administracyjnego (Galligan, Matczak 2005). Godną wspomnienia jest również praca nad pewnymi aspektami orzecznictwa Sądu Najwyższego (Wyrembak 2009).

Założenia metodologiczne badania zostały dostosowane do opisanych celów poznawczych. Badanie miało charakter jakościowy w tym względzie, że zakładało swobodne kodowanie tekstu uzasadnień orzeczeń Trybunału Konstytucyjnego zgodnie z jego interpretacją przez każdego z koderów⁶. Wynikał on również z przyjętej metody generowania kodów użytych w badaniu. Choć nie można uznać jej za ugruntowaną w ścisłym rozumieniu tego słowa, to wychodziła ona z założenia, że wytwarzanie kodów powinno mieć charakter możliwie indukcyjny, wynikać z kontaktu z badanymi

⁶ Wszystkie osoby biorące udział w badaniu ukończyły studia prawnicze, a pięć z nich uzyskało także stopień doktora lub tytuł doktora habilitowanego nauk prawnych. Trzech uczestników badania wykonuje ponadto zawód prawniczy, a trzech ukończyło także studia w zakresie nauk społecznych.

treściami i stanowić przedmiot interakcji pomiędzy badaczami.

Ilościowy charakter badania przejawiał się w założeniu, że treści zaobserwowane w toku analizy powinny być mierzalne, a wynik uzyskany w badaniu próby orzeczeń powinien dać się uogólnić na całą ich populację w drodze wnioskowania statystycznego. Przyjęto także założenie (które odróżnia omawiane badania od podobnych badań realizowanych wcześniej w Polsce), że zbieranie danych należy przeprowadzić przy zachowaniu triangulacji personalnej, to znaczy w drodze równoległego kodowania tego samego materiału badawczego przez więcej niż jednego koderów. Ta ostatnia decyzja wynikała w szczególności ze wspomnianego faktu, że analizowana materia miała charakter szczególnie złożony, a także z wcześniejszych doświadczeń wspomnianych badaczy.

Oprogramowanie

Złożoność i obszerność materiału badawczego, wynikające z powyższych założeń, spowodowały konieczność zastosowania w badaniu wyspecjalizowanego oprogramowania wspomagającego analizę treści. Po rozpoznaniu dostępnych możliwości, zdecydowano się na użycie oprogramowania R w wersji 2.13 (R Core Team 2012) wraz z graficzną nakładką RQDA (nazwa stanowi akronim zwrotu „R-based Qualitative Data Analysis”) w wersji 0.2.2 (Huang 2012). Nakładka ta pozwoliła na łatwe przeprowadzenie kodowania materiału badawczego za pomocą szerokiego zestawu kodów. System statystyczny R umożliwił natomiast dokonanie analizy statystycznej kodów wraz z analizą rzetelności kodowania.

Decyzja o takim doborze oprogramowania była podyktowana kilkoma okolicznościami. Po pierwsze, była to potrzeba dysponowania narzędziem pozwalającym zarówno na swobodne kodowanie złożonych treści za pomocą jakościowo określonych kodów, jak i na wykonanie złożonej analizy statystycznej, w tym także za pomocą procedur, które nie zostały zdefiniowane przez autorów oprogramowania. Po drugie, wynikała ona z faktu, że wymienione narzędzia są dostępne nieodpłatnie – autorzy udostępniają je na licencji GNU GPL – i że można z nich korzystać na kilku systemach operacyjnych. Po trzecie, nie bez znaczenia było to, że program RQDA ma względnie niskie „koszty wejścia”. Ponieważ oferuje użytkownikowi tylko ograniczoną liczbę opcji, pozostawiając wszystkie funkcje analityczne w programie R, jest łatwy w obsłudze i pozwala na udział w badaniach – tylko po krótkim przeszkoleniu – także takim osobom, które nie miały wcześniej do czynienia z tego typu oprogramowaniem⁷.

Program R stanowi jedno z najbardziej rozbudowanych i najbardziej uniwersalnych narzędzi analizy statystycznej dostępnych na komputery PC (jego uruchomienie jest możliwe między innymi na systemach Windows, Linux i MacOSX). Jego historia sięga początku lat dwutysięcznych (Ripley 2001), a program jest wciąż aktywnie rozwijany. Opiera się na zaawansowanym języku programowania, będącym implementacją języka S, używanego w oprogramowaniu komercyjnym⁸, i zawierającym gotowe procedury pozwalające na prowadzenie bardzo

⁷ Warto wspomnieć, że w przypadku nakładki RQDA, oprócz standardowej dokumentacji, nieodpłatnie dostępne są także instrukcje korzystania z programu w formie wideo.

⁸ Pakiet S-PLUS firmy TIBCO. Obydwie implementacje języka S są ze sobą w dużym stopniu kompatybilne.

zaawansowanych analiz statystycznych we wszystkich dziedzinach wiedzy: zarówno w naukach przyrodniczych i technicznych, jak i społecznych czy lingwistyce⁹. Charakterystyczną cechą R – która dla niektórych użytkowników może być wadą – jest to, że w odróżnieniu od innych znanych pakietów statystycznych, takich jak SPSS czy Statistica, nie posiada on interfejsu graficznego. Praca z programem odbywa się natomiast interaktywnie, za pomocą komend wydawanych w wierszu poleceń (środowisku CLI), albo nieinteraktywnie, za pomocą wcześniej przygotowanych skryptów¹⁰.

Istotną własnością R jest także modułowość, przejawiająca się w możliwości instalowania gotowych pakietów rozszerzających standardowy zestaw komend oraz łatwego tworzenia własnych zestawów poleceń. Pakiety rozszerzeń są przygotowywane zarówno przez węższą grupę deweloperów zajmujących się rozwijaniem programu R, jak i przez szerokie kręgi użytkowników, co pozwala na znalezienie gotowych pakietów dopasowanych do rozwiązywania nawet bardzo szczególnych problemów statystycznych. Obecnie (wrzesień 2013) w repozytorium pakietów CRAN, wykorzystywanym przez twórców programu R do dystrybucji oprogramowania, dostępnych jest 4861 takich pa-

⁹ Bardzo obszerna dokumentacja pakietu R jest dostępna na stronie internetowej projektu www.r-project.org (ostatni dostęp: 28 września 2013 r.).

¹⁰ Dostępne jest jednak dodatkowe oprogramowanie ułatwiające korzystanie z R. Są to programy wspomagające tworzenie skryptów w języku R, działające w środowisku graficznym. Istnieje także możliwość instalacji oprogramowania R w środowisku WWW za pomocą odpowiednich nakładek na serwery internetowe, co pozwala na zdalne korzystanie z możliwości oferowanych przez ten pakiet. Wreszcie, istnieją również osobne programy statystyczne dysponujące interfejsem graficznym, porównywalnym z SPSS czy Statistica, i wykorzystujące pakiet R jako „silnik” analiz statystycznych.

kietów. Na szczególną uwagę zasługują praktycznie nieograniczone możliwości wizualizacji danych i szerokie możliwości ich wymiany z innymi programami, oferowane przez oprogramowanie zgromadzone w repozytorium.

Opisywany w niniejszym artykule projekt badawczy wykorzystywał możliwości analizy oferowane przez R jedynie w niewielkim stopniu, ograniczając się do tworzenia tabel kontyngencji, analiz korelacyjnych, analizy logitowej i analiz rzetelności kodowania. Niemniej jednak elastyczność pakietu, przejawiająca się w możliwości łatwego definiowania nowych funkcji i dowolnego łączenia istniejących poleceń oraz w możliwości importowania i swobodnego przekształcania różnorodnych zestawów danych, stanowiła istotny czynnik umożliwiający realizację omawianego badania w formie wynikającej z przyjętych założeń.

Nakładka RQDA, która została wykorzystana w badaniu do bezpośredniego kodowania uzasadnień orzeczeń, stanowi autorski projekt chińskiego socjologa Ronggui Huanga. Do jej uruchomienia niezbędny jest funkcjonujący system R, zaś sama instalacja RQDA odbywa się w sposób zautomatyzowany poprzez wspomniane repozytorium CRAN (program sam pobiera niezbędne komponenty). Interesującą cechą omawianego oprogramowania jest także możliwość wykonania jego tak zwanej statycznej instalacji. Dzięki temu możliwe jest przygotowanie pakietu oprogramowania dla uczestników programu badawczego na nośnikach wymiennych (na przykład pendrive'ach) bez potrzeby jego instalowania na wielu komputerach, co ułatwia realizację badań i pozwala na rozpoczęcie pracy z programem

także użytkownikom o mniejszych kompetencjach z zakresu obsługi komputera¹¹.

Większość danych wytworzonych w nakładce RQDA jest dostępna z poziomu programu R za pomocą specjalnego zestawu komend instalowanych wraz z tą nakładką. W założeniu twórcy, dane wytworzone w programie RQDA powinny podlegać analizie jakościowej lub prostej analizie ilościowej za pomocą przygotowanych przez niego gotowych narzędzi statystycznych, dostępnych z poziomu CLI programu R¹². Ponieważ jednak projekty RQDA zapisywane są jako bazy danych w standardowym formacie programu SQLite, możliwe jest także ich łatwe przetwarzanie za pomocą typowego oprogramowania, w tym wyeksportowanie wytworzonych danych do formatów możliwych do odczytania przez inne programy statystyczne (np. formatu .csv). Możliwe jest również wczytanie danych z nakładki RQDA do systemu R z pominięciem standardowych procedur zaimplementowanych w samym RQDA.

Ta ostatnia możliwość była wykorzystywana w analizie danych wytworzonych w ramach omawianego projektu. Wynika to z faktu, że niezależnie od wielu możliwości analizy oferowanych przez oprogramowanie RQDA, nie jest ono bezpośrednio przystosowane do przeprowadzania badań wykorzystujących triangulację personalną i nie zawiera opcji równoległej pracy wielu koderów na tym sa-

¹¹ Ta możliwość jest z zasady wykluczona w przypadku narzędzi badawczych udostępnianych na zasadach komercyjnych, w przypadku których liczba dokonywanych instalacji jest ściśle ograniczona do komputerów objętych licencją.

¹² Ich dokładny opis jest zawarty w dokumentacji programu, dostępnej na jego stronie internetowej <http://rqda.r-forge.r-project.org/> (ostatni dostęp: 28 września 2013 r.).

mych materiale¹³. Z tego względu w badaniu wykorzystano pakiet dodatkowy „RSQLite” pozwalający na odczyt w programie R baz danych zapisanych w formacie SQLite. Za jego pomocą połączono zbiory danych wytworzone przez poszczególnych koderów, zachowując informację o autorstwie poszczególnych zapisów.

Oprogramowanie RQDA zostało zaprojektowane w taki sposób, aby pozwalało na kodowanie swobodne, to jest bez narzuconej z góry jednostki (czy segmentu) analizy. Kodowanie może się odbywać zarówno za pomocą kodów wytwarzanych *in vivo* przez samego koder, jak i określonych wcześniej. Dzięki temu, a także za sprawą opcji grupowania kodów, tworzenia notatek i dzienników pracy możliwe jest prowadzenie za pomocą RQDA analiz o charakterze czysto jakościowym, w tym odwołujących się zwłaszcza do wytycznych metody ugruntowanej. Oprogramowanie RQDA zostało ponadto przystosowane do wytwarzania danych niezbędnych do analizy statystycznej typu QCA (*qualitative comparative analysis*, dalej QCA), to jest nieprobabilistycznej analizy danych ilościowych o niewielkiej liczbie obserwacji. Możliwe jest także opisywanie poszczególnych plików za pomocą dowolnej liczby

¹³ Istnieje natomiast opcja pracy na tym samym materiale bez stosowania triangulacji. W takim przypadku możliwe jest wykorzystanie zdalnej bazy danych MySQL umieszczonej na serwerze dostępnym z sieci Internet, z którą łączą się badacze korzystający z programu RQDA na swoich lokalnych komputerach. Takie rozwiązanie, w połączeniu ze wspomnianymi wyżej nakładkami instalującymi R na serwerach WWW, może (po przewyżczeniu dodatkowych trudności technicznych) pozwalać na automatyczne informowanie badaczy o wynikach badania w trakcie kodowania, co może przyczynić się do wzrostu refleksyjności całego procesu. Warto zauważyć, że w takiej konfiguracji omawiane oprogramowanie nie narzucałoby praktycznie żadnych limitów w sprawie liczby koderów biorących udział w projekcie (ograniczeniem byłaby tu jedynie pojemność i szybkość działania bazy danych).

zmiennych numerycznych lub tekstowych, dających się następnie wykorzystać w klasycznej analizie ilościowej, na przykład jako zmienne niezależne.

Interfejs graficzny programu RQDA można opisać jako nieskomplikowany i przez to intuicyjny. Grupyje on opcje dostępne dla użytkownika w osiem kategorii: menu „Project”, „Files” i „Settings” pozwalają na zarządzanie projektami i dodawanie do nich nowych plików do zakodowania. RQDA pozwala tu na importowanie jedynie plików tekstowych, a więc pliki innego rodzaju muszą zostać przekształcone na ten format przed rozpoczęciem kodowania¹⁴. Menu „Codes” pozwala na definiowanie kodów wykorzystywanych do kodowania zaimportowanych plików, wyświetlanie fragmentów tekstu zakodowanych za pomocą poszczególnych kodów oraz tworzenie notatek odnoszących się do konkretnych fragmentów kodowanego materiału. Samo kodowanie odbywa się poprzez zaznaczenie wybranego fragmentu tekstu i wybór odpowiedniego kodu z menu programu¹⁵. Menu „Code categories” zawiera opcje ułatwiające grupowanie kodów, pozwalając tym samym na budowanie teorii w ramach ugruntowanych strategii badawczych lub w celu uporządkowania zebranych danych. Ograniczeniem RQDA w tym zakresie jest niemożność tworzenia wielopoziomowych kategoryzacji kodów – program pozwala jedynie na grupowanie kodów za pomocą kategorii mieszczących się na tym samym poziomie klasyfikacji, a więc nie daje możliwości określenia ich hierarchii lub innych relacji. Menu „Cases” pozwala na pogrupowanie analizowanych plików w sposób dogodny dla użyt-

kownika, na przykład połączenie ich w jedną jednostkę analizy na potrzeby badania QCA. Menu „Attributes” pozwala na zdefiniowanie zmiennych opisujących poszczególne pliki lub ich grupy (*cases*). Menu „File categories” pozwala na grupowanie plików w kategorie. Wreszcie, menu „Journals” umożliwia tworzenie notatek z prowadzonych prac z uwzględnieniem chronologii. Jak to zostało powiedziane, RQDA nie zawiera natomiast żadnych bardziej zaawansowanych opcji analizy danych dostępnych z poziomu interfejsu graficznego. Zgodnie z założeniem autora programu, ilościowa analiza danych ma być prowadzona z poziomu programu R. Podobnie jest z opcją wyszukiwania określonych fragmentów tekstu¹⁶.

Najważniejszą cechą RQDA jako narzędzia analizy treści jest to, że pozwala ono na prowadzenie całkowicie swobodnego kodowania analizowanych tekstów. Badacz oznaczający jakiś fragment tekstu wybranym przez siebie kodem może uczynić to w odniesieniu do fragmentu tekstu o dowolnej długości, poczynając od 1 znaku. Możliwe jest przy tym jednocześnie (wielokrotne) zakodowanie tego samego fragmentu tekstu dowolną liczbą kodów. Zakodowane fragmenty mogą też pozostawać w dowolnych relacjach: pokrywania się, zawierania, krzyżowania lub rozłączności.

¹⁶ Elastyczność oprogramowania R pozwala na jego wykorzystanie do zautomatyzowanego lub półautomatycznego kodowania analizowanych tekstów w oparciu o ich zadane cechy (na przykład przy wykorzystaniu wyszukiwań pełnotekstowych). W omawianym badaniu tej opcji jednak nie wykorzystywano; można sądzić, że wymagałoby to stworzenia własnej sekwencji komend w języku R. Autor oprogramowania RQDA przygotował również pakiet dodatkowy RQDA_{tm}, służący do realizacji badań ilościowych na korpusie wypowiedzi. Rozszerza on możliwości R i RQDA o funkcje typowe dla badań lingwistycznych (analiza kolokacji, konkordancji itp.), jego ograniczeniem może być jednak niedostosowanie do badania tekstów w językach innych niż angielski.

Wewnętrzny format zapisu danych wykorzystywany w programie RQDA zapewnia użytkownikowi prowadzącemu analizę dostęp do wszystkich danych wytworzonych podczas kodowania. W szczególności z poziomu programu R (lub innego programu statystycznego) możliwe jest odczytanie następujących informacji umieszczonych w bazie danych: kod użyty do oznaczenia danego fragmentu tekstu, plik, w którym znajduje się zakodowany fragment, początek zakodowanego fragmentu jako odległość w znakach od początku kodowanego tekstu, koniec zakodowanego fragmentu (analogicznie), zakodowany tekst, wartości zmiennych (atrybutów) ustalonych dla poszczególnych plików. Pozwala to na zachowanie pełnej wiedzy o zakodowanych fragmentach tekstu także podczas analizy ilościowej i również po wyeksportowaniu danych stworzonych w RQDA do innego niż R programu statystycznego.

Metodologia badania orzecznictwa Trybunału Konstytucyjnego

Opisane cechy oprogramowania R i RQDA pozwoliły na przeprowadzenie badań uzasadnień orzeczeń Trybunału Konstytucyjnego w sposób zgodny z potrzebami poznawczymi, posiadanymi zasobami i założeniami wyjściowymi. Własności narzędzia informatycznego nie nałożyły zwłaszcza większych ograniczeń na konceptualizację badań. Zastosowana procedura badawcza była kilkuetapowa. W pierwszej kolejności przeprowadzono pilotaż polegający na swobodnym kodowaniu – przez wszystkich sześciu badaczy biorących udział w projekcie – pięciu dobranych celowo orzeczeń Trybunału Konstytucyjnego. Na tym etapie ba-

dania każdy z koderów pracował na identycznym materiale, przeprowadzając jego niezależne kodowanie za pomocą samodzielnie wytwarzanych kodów *in vivo* – w trakcie lektury kodowanych orzeczeń. Jedynym ustalonym *explicite* ograniczeniem dla inwencji badaczy było tu uzgodnione założenie, że celem badania jest ustalenie wzorców argumentacji wykorzystywanych przez Trybunał.

W ten sposób wytworzono 312 kodów o przecinających się zakresach i różnym stopniu wykorzystania przez różnych koderów i w różnych orzeczeniach. Następnie przeprowadzono kilka sesji uzgodnieniowych zmierzających do ustalenia listy kodów wspólnej dla zespołu badawczego oraz do ujednoczenia znaczeń przypisywanych poszczególnym kodom przez poszczególnych badaczy.

W efekcie tych działań wytworzono listę 50 kodów, z których większość opatrzone notatkami opisującymi ich uzgodnioną treść. Kody ustalone na tym etapie były dwójakiego rodzaju. Kody pierwszego poziomu (w liczbie 37) opisywały argumenty wysuwane przez Trybunał Konstytucyjny w uzasadnieniach orzeczeń. Kody drugiego poziomu (13) służyły do tego, by scharakteryzować rodzaj tekstu prawnego, do którego odnosiły się kody pierwszego poziomu. Rozróżnienie to miało służyć przede wszystkim ustaleniu, czy i w jakim stopniu wzorce argumentacji stosowane w odniesieniu do różnych aktów prawnych różnią się od siebie¹⁷. Listę kodów przedstawia Aneks.

¹⁷ W konsekwencji każdy fragment uzasadnienia orzeczenia TK, odnoszący się do jakiegoś aktu prawnego, był kodowany co najmniej dwukrotnie – przynajmniej raz kodem pierwszego poziomu i raz kodem drugiego poziomu.

¹⁴ Program obsługuje polskie znaki diakrytyczne w formacie utf8.

¹⁵ Program dopuszcza przy tym edytowanie zawartości kodowanego pliku bez straty kodowania.

Jego lektura pozwoli łatwo zauważyć, że niezależnie od zaakceptowanej przez wszystkich badaczy indukcyjnej metody wytwarzania kodów – zgodnie z którą mogły one być określone w sposób dowolny, ale przede wszystkim na podstawie lektury uzasadnień orzeczeń – dominują wśród nich kategorie pojęciowe dobrze znane teorii prawa i wykorzystywane w dyskursie prawniczym. Ich status metodologiczny jest też zróżnicowany – są to zarówno kody odnoszące się do tradycyjnych, znanych od setek lat rozumowań czy topik prawniczych, jak i do koncepcji wypracowanych w ramach konkretnych doktryn wykładni. Sprawilo to, że typologia sposobów argumentowania, wyrażona za pomocą kodów, nie była skonstruowana według jednolitego schematu i w konsekwencji nie była logicznie zupełna. Niektóre jej elementy nie były też rozłączne, a granice ich zakresów były niekiedy nieostre.

Na drugim etapie badania zrealizowano główną część analizy treści. W jej trakcie – za pomocą ustalonych kodów – przeprowadzono badanie 150 uzasadnień orzeczeń TK z lat 1986–2009¹⁸. Orzeczenia zostały dobrane w taki sposób, aby były reprezentatywne w dwóch wymiarach. Po pierwsze, chodziło o dwa istotne okresy w historii działalności TK w Polsce, to znaczy czas przed 1997 rokiem (kiedy obowiązywały poprzednie przepisy konstytucyjne) i po 1997 roku (kiedy w życie weszła nowa konstytucja). Po drugie,

¹⁸ Według bazy danych „Orzecznictwo Trybunału Konstytucyjnego 1986–2009” wydanej przez Biuro Trybunału Konstytucyjnego z datą 3 maja 2010 r., w badanym okresie TK wydał łącznie 5521 orzeczeń, z czego 1908 rozstrzygały problem prawny co do meritum (pozostałe orzeczenia nosiły sygnatury rozpoczynające się od „T”, a więc dotyczyły tzw. wstępnej kontroli dopuszczalności orzekania). Za operat stanowiący podstawę doboru próby przyjęto orzeczenia należące do tej pierwszej kategorii. Dobór orzeczeń do próby w ramach ustalonych kategorii miał charakter losowy.

złożono reprezentatywność ze względu na rodzaj orzeczenia określony jego sygnaturą¹⁹. Tak dobrane orzeczenia poddano zautomatyzowanej obróbce za pomocą prostego narzędzia informatycznego, usuwając z nich wszystkie elementy poza uzasadnieniem stanowiska TK²⁰. Łączna objętość tekstów poddanych badaniu (po obróbce) wynosiła 3 604 822 znaków (średnio 24 032 znaki na uzasadnienie), co stanowi 7,2% całego korpusu orzeczeń Trybunału po obróbce i 4,2% przed jej dokonaniem.

Na tym etapie badania kody nie podlegały już dalszym modyfikacjom. Badacze nie dodawali również nowych, a jedynie kodowali uzasadnienia orzeczeń w sposób swobodny, bez zdefiniowanej jednostki analizy i z wykorzystaniem możliwości wielokrotnego kodowania tego samego fragmentu tekstu różnymi kodami. Ta ostatnia możliwość wynikała ze wspomnianego faktu, że nie wszystkie kody były

¹⁹ Orzeczenia TK są oznaczane następującymi sygnaturami: K – orzeczenie odnoszące się do niezgodności z konstytucją przepisów ustaw, Sk – orzeczenie wydane w efekcie skargi konstytucyjnej, P – orzeczenie wydane w efekcie pytania prawnego sądu, U – orzeczenie odnoszące się do niezgodności z przepisami rangi ustawowej lub konstytucją aktów podstawowych (rozporządzeń), S – wyroki sygnalizujące potrzebę nowelizacji prawa, W – wyroki wypowiedziane powszechnie wiążącą interpretacją prawa (do 1997 roku).

²⁰ Orzeczenia TK kończące sprawę co do meritum składają się najczęściej z czterech części. Część pierwsza to sentencja wyroku, w której Trybunał orzeka o uchyleniu lub utrzymaniu w mocy konkretnych przepisów prawa. W tej części zawarta jest również informacja o składzie sędziowskim, który wydał wyrok, stronach postępowania itd. Kolejne trzy części składają się na uzasadnienie wyroku. W części pierwszej uzasadnienia Trybunał przedstawia pisemne stanowisko strony, która przedstawiła wniosek, pytanie prawne lub skargę konstytucyjną oraz stanowiska innych podmiotów. W części drugiej zawarte jest streszczenie stanowisk podmiotów biorących udział w postępowaniu, przedstawione na rozprawie. Dopiero trzecia część uzasadnienia zawiera autorskie wypowiedzi Trybunału. Przedmiotem badania była tylko owa trzecia część, co wynika z faktu, że pozostałe elementy uzasadnienia stanowią zazwyczaj mechaniczne powtórzenie treści zawartych w piśmie przedstawianych przez strony postępowania. Gdy dalej w niniejszym tekście jest mowa o „uzasadnieniu orzeczenia”, chodzi właśnie o tę część.

w swoich zakresach całkowicie rozłączne i same podlegały interpretacji, z faktu, że stosowano kody znajdujące się na dwóch poziomach, jak i faktu, że wypowiedzi TK były złożone i niekiedy nie dawały się opisywać za pomocą tylko jednego kodu.

Zastosowano ponadto triangulację personalną. Polegała ona na podzieleniu 150 uzasadnień orzeczeń pomiędzy członków zespołu badawczego w taki sposób, aby każde orzeczenie zostało zakodowane przez dwóch badaczy i jednocześnie, by poszczególne pary badaczy powtarzały się z równą częstością. W efekcie, każdy z koderów analizował 50 orzeczeń, z czego każde 10 było analizowane również przez innego z pozostałych pięciu koderów. W ten sposób uniknięto tworzenia „stałych par” koderów, co mogłoby się przyczynić do zmniejszenia intersubiektywnej sprawdzalności wyniku.

W efekcie przeprowadzonych działań badawczych przypisano kody do 15 080 fragmentów uzasadnień orzeczeń (tj. do około 50 fragmentów na każde uzasadnienie na jednego kodera). Objętość zakodowanych fragmentów tekstu wynosiła 14 896 778 znaków, co oznacza, że wybrane do badania uzasadnienia orzeczeń zostały pokryte kodami przez każdego kodera średnio około dwukrotnie (tj. każdy znak w każdym uzasadnieniu był przypisany do co najmniej dwóch fragmentów tekstu określonych przez kodera i oznaczonych różnymi kodami). Pozwala to sądzić, że przeprowadzona analiza miała charakter kompletny, to znaczy objęła całość materiałów przeznaczonych do badania, ale jednocześnie, że możliwość wielokrotnego kodowania tych samych fragmentów tekstu różnymi kodami przez tego samego kodera nie była często wykorzystywana.

Na tym etapie badania do zbioru danych dołączono też informacje o poszczególnych orzeczeniach, zakodowane jako atrybuty poszczególnych plików. Były to następujące zmienne: sygnatura orzeczenia, rok jego wydania, rodzaj rozstrzygnięcia Trybunału (zgodność, niezgodność, brak niezgodności poszczególnych przepisów z konstytucją lub innymi aktami albo umorzenie postępowania)²¹, liczba zdań odrębnych, wielkość składu sędziowskiego. Dodatkowo dokonano syntetycznej oceny zgodności rozstrzygnięcia Trybunału z przypuszczalnymi celami przedłożonego mu wniosku, kodując zgodność lub niezgodność orzeczenia z żądaniem wnioskodawcy za pomocą zmiennej zerojedynkowej²². Podczas analizy informacje te posłużyły jako zmienne wyjaśniające zaobserwowaną zmienność wzorców argumentacji TK.

Treści wygenerowane w efekcie opisanej procedury nie były poddawane dalszej obróbce czy modyfikacjom, a ich analiza – stanowiąca trzecią część procedury badawczej – miała charakter ściśle ilościowy. W szczególności, na podstawie tak wygenerowanych kodów, nie budowano teorii w rozumieniu metodologii ugruntowanej. Nie próbowano też uzgadniać sposobu kodowania wybranych fragmentów pomiędzy badaczami. Przeciwno takim działaniom przemawiała znaczna objętość kodowanego materiału, rezygnacja z kodowania *in vivo* podczas zasadniczego etapu analizy, a także doświadczenia związane z sesjami uzgodnieniowymi

²¹ Ponieważ w każdym orzeczeniu możliwe jest wydanie dowolnej kombinacji decyzji w odniesieniu do poszczególnych przepisów stanowiących przedmiot orzekania, w badaniu kodowano tę kwestię jako liczbę rozstrzygnięć poszczególnego rodzaju w danym orzeczeniu.

²² Tę część prac badawczych przeprowadził student WPIA UW, pan Jędrzej Maśnicki.

na etapie pilotażowym, które dowiodły znacznej pracochłonności takich działań w sześciuosobowej grupie badaczy.

Dla celów analizy statystycznej wykorzystano dwie syntetyczne miary: względną objętość kodu w całym korpusie uzasadnień orzeczeń oraz jego częstość. Ta pierwsza zmienna została obliczona jako stosunek sumy objętości wszystkich fragmentów zakodowanych danym kodem (w znakach) do sumy objętości wszystkich fragmentów zakodowanych wszystkimi kodami danego poziomu. Wartości tej drugiej ustalono jako stosunek liczby użyć danego kodu we wszystkich orzeczeniach do sumy użyć wszystkich kodów we wszystkich orzeczeniach. Tak zdefiniowane miary posłużyły do stworzenia tabeli kontyngencji poszczególnych kodów i analiz korelacyjnych (za pomocą miary właściwej dla skal nominalnych, tj. V Cramera) oraz (po przeprowadzeniu stosownego przekształcenia kodów ze skali nominalnej na zerojedynkową) także do przeprowadzenia analizy logitowej. Dla ustalenia istotności wyników zawartych w tabelach kontyngencji użyto testu χ^2 ²³. Jego zastosowanie dowiodło, że wszystkie ustalenia powzięte w toku analizy były istotne statystycznie, zazwyczaj na poziomach istotności $p < 0,01$. Przeprowadzono także analizę rozkładu kodów (tj. zróżnicowania korpusu ze względu na kody w poszczególnych orzeczeniach) oraz współwystępowania kodów pierwszego poziomu.

Warto zauważyć, że pewną komplikację w analizie danych i prezentacji wyników stanowi konieczność

²³ Analizę istotności przeprowadzono dla bezwzględnych wartości obydwu wspomnianych miar.

uwzględnienia w nich triangulacji personalnej. Ponieważ każde z analizowanych uzasadnień orzeczeń było kodowane niezależnie przez dwóch koderów, niecelowe było podawanie wyników badania w miarach bezwzględnych (tj. liczby wystąpień poszczególnych kodów w korpusie). Takie działanie znacznie zawyżyłoby rzeczywiste występowanie poszczególnych kodów w korpusie, prowadząc do błędnych wniosków.

Rozwiązaniem tego problemu może być podawanie wspomnianych miar względnych. Takie działanie może być jednak obciążone błędem wynikającym z różnic w sposobie kodowania tego samego materiału przez różnych koderów, zwłaszcza związanych z różną szczegółowością czy „gęstością” kodowania. Może się mianowicie zdarzyć, że jeden z pary koderów ma tendencję do wyodrębniania dłuższych segmentów tekstu, zaś drugi do dzielenia go na krótsze części. W takiej sytuacji (nawet gdy w parze koderów zostanie zachowana całkowita zgodność w kwestii kodu użytego do oznaczenia danego fragmentu) syntetyczna miara częstości kodów będzie w większym stopniu odzwierciedlać sposób kodowania drugiego z koderów, zaś miara długości – pierwszego. Z tego względu w badaniu wykorzystano proste wagi analityczne, zmniejszające wpływ stylu kodowania poszczególnych koderów na wynik. Zostały one ustalone dla każdego analizowanego uzasadnienia, osobno dla miar względnej objętości i częstości. Wagi te obliczono jako różnicę między liczbą 1 a ilorazem całkowitej objętości (liczebności) fragmentów tekstu zakodowanych w danym orzeczeniu przez danego koderów oraz sumą objętości (liczebności) fragmentów tekstu zakodowanych w danym orzeczeniu przez obydwu koderów, którzy je

kodowali. Porównanie wyników uzyskanych za pomocą danych ważonych i nieważonych nie prowadziło jednak do zmiany zasadniczych wniosków płynących z badania. Można to w jakimś stopniu wiązać z zastosowaną techniką triangulacji, przewidującą brak stałych par koderów.

Zapewne najbardziej interesujące zagadnienie analityczno-metodologiczne, które pojawiło się podczas analizy danych, wiąże się z kwestią rzetelności kodowania (*intercoder reliability*). Zastosowanie triangulacji personalnej w naturalny sposób prowadzi mianowicie do potrzeby oceny zgodności pomiędzy sposobem kodowania tych samych fragmentów tekstu przez różnych koderów (Neuendorf 2002: 141 i nast.; Krippendorff 2004a: 211 i nast.). W systemie R obliczenie statystycznych miar rzetelności nie naraża większego problemu, ponieważ dostępne pakiety dodatkowe dostarczają gotowych procedur pozwalających wyznaczyć rozmaite współczynniki, takie jak Kappa Cohena, współczynnik ICC czy Alfa Krippendorfa (w badaniu użyto pakietu dodatkowego „irr”). W badaniu zdecydowano się wykorzystać, jako posiadającą najbardziej korzystne własności matematyczne, Alfę Krippendorfa (Krippendorff 2004b).

Wszystkie jednak miary tego rodzaju, znane autorowi analizy – i wszystkie zaimplementowane w środowisku R – zakładają porównanie kodów zastosowanych do określonej z góry jednostki analizy, to znaczy opierają się na porównaniu kodów użytych przez różnych koderów do oznaczenia tych samych fragmentów tekstu. Jak to zostało powiedziane, w przypadku omawianych badań taka jednostka analizy nie istniała. Każdy z pary koderów kodu-

jących to samo orzeczenie mógł zastosować każdy z 50 kodów na oznaczenie fragmentów tekstu pozostających w dowolnych relacjach, co uniemożliwiało bezpośredni pomiar rzetelności.

Literatura odnosząca się do pomiaru rzetelności kodowania (na przykład Carey, Morgan, Oxtoby 1996; Lombard, Snyder-Duch, Campanella Bracken 2002; Krippendorff 2004a: 211 i nast.; 2004b; 2008; Carey i in. 2004 i tam cytowane prace) nie zawiera wielu propozycji rozwiązania tego problemu. Jedną z możliwości stanowi oddzielenie segmentacji tekstu od przypisywania kodów do segmentów (Campbell i in. 2013), to znaczy zastosowanie procedury, w której jeden z badaczy przeprowadza pełne, swobodne kodowanie, a pozostali jedynie kodują wskazane przez niego segmenty, ale nie biorą udziału w ich definiowaniu. Wadą tego rozwiązania jest to, że nie pozwala ono na dokonanie triangulacji samego schematu segmentacji, a więc to, iż w tym względzie w istocie nie różni się od sytuacji, gdy jednostka analizy jest z góry ustalona.

Inna autorka (Kurasaki 2000) proponuje stosowanie w takiej sytuacji losowego próbkowania porównywanych fragmentów tekstu i stosowania miar rzetelności tylko do nich. Rozwiązanie to, choć jest niewątpliwie eleganckie i pozwala na ominięcie problemu segmentacji kodowanego tekstu, jest także obciążone pewnymi wadami. Powoduje zwłaszcza, że zagadnienie rzetelności kodowania nabywa charakteru podwójnie probabilistycznego²⁴.

²⁴ Już w sytuacji kompletnego pomiaru rzetelności kodowania w badaniach na materiale stanowiącym próbę pojawia się problem wnioskowania na temat (hipotetycznej) rzetelności kodowania w całej populacji. Zastosowanie próbkowania będzie, rzecz jasna, obniżyć ufność pomiaru rzetelności.

Prowadzi to do dalszych komplikacji, takich jak na przykład wpływ wielkości próbkowanego fragmentu i liczby próbek na istotność oszacowania rzetelności. To z kolei wpływa na pewność wyników całego badania, które same mają charakter probabilistyczny (można twierdzić, że za miarę ufności badania należałoby przyjąć iloczyn poziomu ufności wyniku i poziomu ufności rzetelności kodowania)²⁵.

Nie chcąc rozstrzygać takich problemów – jako wtórnych wobec głównego tematu badań – zastosowano rozwiązanie polegające na wtórnej (sztucznej) segmentacji już zakodowanego tekstu. Ponieważ rozwiązanie takie nie jest często spotykane, przeprowadzono swojego rodzaju eksperyment metodologiczny, polegający na dokonaniu pomiaru rzetelności kodowania przy wykorzystaniu kilku różnych schematów segmentacji tekstu i zestawieniu ze sobą uzyskanych wyników.

W pierwszym wariantcie tej analizy za segment tekstu poddawany porównaniu w parze koderów, którzy go kodowali, uznano całe uzasadnienie orzeczenia. Działanie to opierało się na przeświadczeniu, że można opisać dane uzasadnienie za pomocą wartości objętości względnej i częstości każdego z kodów użytych w tym orzeczeniu, traktując owe wartości jako „zmierzoną” przez *i*-tego koderów „intensywność” występowania tego kodu w danym

uzasadnieniu²⁶. Dokonując takich obliczeń dla każdego koderów, kodu i uzasadnienia, uzyskano tabelę zawierającą 15 000 *data points* (50 orzeczeń * 50 kodów * 6 koderów). Następnie porównano „intensywność” występowania poszczególnych kodów w tych samych orzeczeniach kodowanych przez poszczególne pary koderów, za stan idealnej zgodności kodowania uznając taką sytuację, w której „intensywność” występowania danego kodu w danym orzeczeniu byłaby taka sama według *i*-tego i *j*-tego koderów kodującego to orzeczenie²⁷. Jako miarę rzetelności wykorzystano Alfę Krippendorfa w wariantcie przewidzianym dla skali ilorazowej.

Drugi wariant segmentacji polegał na tym, że analizę przeprowadzono na arbitralnie wydzielonych 1000-znakowych fragmentach uzasadnień orzeczeń, którym przypisano występujące w nich kody. W przypadku, gdy w danym fragmencie występował więcej niż jeden kod, traktowano je jak dwa lub więcej osobne fragmenty (nieliczne fragmenty niezakodowane żadnym kodem wyłączało z analizy). Następnie dokonano porównania wszystkich odpowiadających sobie par fragmentów uzasadnień orzeczeń, zakodowanych przez różnych koderów, stwierdzając, czy występuje pomiędzy nimi zgodność zastosowanego kodu i czy w ogóle obydwa koderzy zakodowali ten fragment jakimkolwiek kodem. Na tej podstawie obliczono Alfę Krippendorfa dla każdego kodu i każdego koderów (porów-

²⁵ Teoretycznym rozwiązaniem może być tu zastosowanie „pełnego próbkowania”, to znaczy porównywanie kodów zastosowanych przez różnych koderów w tym samym tekście dla każdego znaku. Rozwiązanie takie będzie – gdy chodzi o stosowany algorytm – zbliżone do drugiej z metod segmentacji, opisanych poniżej. Jego wadą jest jednak to, że prowadzi do wygenerowania dużej ilości danych (trójwymiarowej tabeli o rozmiarach $x = \text{długość porównywanego tekstu}$, $y = \text{liczba kodów}$, $z = \text{liczba koderów}$).

²⁶ Z zastrzeżeniem, że suma wartości zarówno objętości względnej, jak i częstości dla wszystkich kodów mogła dla każdego orzeczenia wynosić więcej niż 1 ze względu na wielokrotne kodowanie.

²⁷ Trzeba podkreślić, że tego rodzaju procedura opiera się na porównaniu zgodności wartości intensywności kodów dla danego orzeczenia, nie zaś współwystępowania samych kodów.

nując sposób kodowania *i*-tego koderów i wszystkich pozostałych).

W trzecim wariantcie pomiaru rzetelności przyjęto, że sposób kodowania porównuje się pomiędzy wszystkimi fragmentami tekstu pochodzącymi z tego samego uzasadnienia i zakodowanymi przez różnych koderów, jeśli fragmenty te pozostają w jednej z następujących relacji: pokrywanie się (fragmenty zakodowane przez różnych koderów rozpoczynają się w tym samym miejscu orzeczenia i kończą się w tym samym miejscu), zawieranie się (jeden z fragmentów w całości obejmuje drugi), przecinanie się (jeden z fragmentów rozpoczyna się w miejscu, które zawiera tekst drugiego) lub bliskość (tekst obydwu fragmentów jest rozłączny, ale początek jednego z nich znajduje się w odległości nie mniejszej niż 50 znaków od końca drugiego)²⁸. W tym celu wykorzystano Alfę Krippendorfa w wariantcie przewidzianym dla skali nominalnej.

Konsekwencją zastosowania dwóch pierwszych z wymienionych schematów segmentacji tekstu i metod obliczania rzetelności było to, że porównanie rzetelności mogło odnosić się jedynie do poszczególnych kodów, nie zaś do wszystkich zakodowanych fragmentów. Tylko zastosowanie trzeciego schematu segmentacji pozwoliło na obliczenie całościowej rzetelności kodowania dla fragmentów tekstu zakodo-

²⁸ Warto zauważyć, że w języku R implementacja takiego sposobu segmentacji może okazać się bardzo zasobochłonna i w efekcie powolna. Najprostszy (z punktu widzenia standardowej implementacji informatycznej) algorytm segmentacji, który został zastosowany w omawianych badaniach, opierający się na zagnieżdżonych pętach *for*, okazał się bardzo powolny (jego wykonanie na danych z badania trwało kilka godzin na komputerze z czterordzeniowym procesorem) i powinien zostać przebudowany w taki sposób, aby wykorzystywał działania na wektorach. Ma to związek z ogólniejszą architekturą systemu R, która zakłada właśnie taki mechanizm operowania na danych.

wanych przez *i*-tego koderów i wszystkich pozostałych koderów, i wszystkich kodów.

Uzyskane w efekcie każdej z procedur wartości Alfę Krippendorfa odbiegały od poziomów uznawanych za satysfakcjonujące – 0,8 i wystarczające – 0,67 (por. Krippendorff 2004a: 241; 2008). W przypadku pierwszych dwóch mechanizmów segmentacji tekstu, uzyskane miary były też bardzo zróżnicowane i wyraźnie zależne od częstości użycia i objętości poszczególnych kodów. W przypadku niektórych kodów i niektórych koderów występowały ujemne wartości Alfę Krippendorfa. W przypadku innych kodów uzyskano wartości przekraczające 0,67. Zaobserwowano również takie sytuacje, w których niektórzy z koderów uzyskiwali niskie lub bardzo niskie wartości Alfę, zaś inni, w odniesieniu do tych samych kodów – wysokie lub bardzo wysokie. Gdy wreszcie chodzi o wartości uzyskane w efekcie trzeciego z opisanych algorytmów, to pozostawały one w przedziale 0,29–0,41.

Wyniki te można interpretować na kilka sposobów. Po pierwsze, mogą one świadczyć o rzeczywistych różnicach w interpretacji tych samych fragmentów tekstu przez różnych koderów. Wydaje się to prawdopodobne zwłaszcza w świetle znacznego poziomu skomplikowania kodowanego tekstu. Po drugie, można sądzić, że miary te stanowią w pewnej mierze artefakt wynikający ze względnie niskiej częstości niektórych kodów (ich wąskiego zakresu) lub sposobu przeprowadzenia obliczeń rzetelności. Po trzecie, można uważać, że mogą być one związane z nieostrością samych kodów lub różnicami w ich interpretacji.

Dla poprawy rzetelności wyników badania i choć częściowego wyjaśnienia tych wątpliwości, w toku

dalszej analizie zdecydowano się na przeprowadzenie agregacji kodów (to jest połączenie ze sobą wybranych kodów w kody o szerszym zakresie i w konsekwencji objętości zakodowanego tekstu i częstości wystąpień równej sumie objętości i wystąpień kodów „składowych”). Ponieważ wynik tego działania zależy od przyjętego schematu agregacji, który z kolei może być zależny od indywidualnych preferencji osoby definiującej ten schemat, agregację zrealizowano zgodnie ze specjalnym algorytmem.

Przedstawiał się on następująco. W pierwszej kolejności, każdy z sześciu koderów biorących udział w badaniu przedstawił własny schemat agregacji kodów pierwszego poziomu. Następnie, dla każdego zaproponowanego schematu agregacji przeprowadzono analizę rzetelności kodowania za pomocą trzeciego z wymienionych wyżej algorytmów, modyfikując przy tym założenie dotyczące warunku bliskości kodów, tak aby uwzględnił trzy warianty. Odległość pomiędzy porównywanymi fragmentami wynosiła w nich odpowiednio 5, 50 i 1000 znaków²⁹. Wreszcie, jako kryterium wyboru docelowego schematu agregacji przyjęto wartości Alfę we wszystkich trzech wariantach, po uśrednieniu wyników uzyskanych dla wszystkich sześciu koderów objętych badaniem.

²⁹ Odległość między fragmentami może mieć znaczenie dla wartości Alfę Krippendorfa, bo im większa jest akceptowana odległość, tym więcej fragmentów tekstu jest porównywanych pod względem sposobu zakodowania. W konsekwencji, z jednej strony, wraz ze wzrostem akceptowanej odległości wzrasta liczba zaobserwowanych niezgodności pomiędzy parami kodów zastosowanych przez różnych koderów, ale też zwiększa się prawdopodobieństwo wystąpienia par zgodnych. Opisany dobór odległości stanowi swojego rodzaju eksperyment metodologiczny, związany z nieostrością wypowiedzi podlegających kodowaniu i nieostrością samych zastosowanych kategorii. W szczególności przyjęcie odległości 1000-znakowej może wydawać się przesadne, jednak opierało się na przeświadczeniu, że dostrzeżone przez jednego z koderów aspekty wypowiedzi TK mogą przez drugiego z nich zostać zaobserwowane w innych miejscach tekstu, w tym dość odległych.

Zgodnie z oczekiwaniami, agregacja przyniosła istotną poprawę wskaźników rzetelności kodowania. W najlepszym z badanych schematów agregacji Alfa Krippendorfa wynosiła od 0,5 do 0,7, w zależności od koderów. Ostatecznie zredukowano liczbę kodów pierwszego poziomu z 37 do 12, a kodów drugiego poziomu – z 13 do 3. Można sądzić, że taka procedura agregacji stanowi w istocie swoisty odpowiednik sesji uzgodnieniowych i, w pewnym zakresie, funkcjonalny ekwiwalent budowania teorii ugruntowanej w drodze tworzenia grup kodów³⁰.

CAQDA a socjologiczna analiza orzecznictwa sądowego

Niezależnie od zasadniczej potrzeby uzyskania wysokich miar rzetelności, warto jednak zauważyć, że wykorzystanie triangulacji personalnej w badaniach opierających się na opisywanym schemacie prowadzi do interesującego dylematu. Nie jest mianowicie zupełnie jasne, czy sytuacja występowania całkowitej czy bardzo wysokiej zgodności pomiędzy koderami jest stanem zawsze pożądanym. Odpowiedź na to pytanie wydaje się oczywista w przypadku pomiaru rzetelności kodowania prostych tekstów (na przykład krótkich odpowiedzi na pytania otwarte w badaniach ankietowych). W takim przypadku interpretacja wypowiedzi badanych przez poszczególnych kode-

³⁰ Jakkolwiek w omawianych badaniach dalsza analiza danych miała charakter ilościowy, otwiera to interesujące możliwości wypracowania komputerowo wspomaganego metodologii jakościowego budowania teorii na podstawie materiału zgromadzonego przez kilkusobowy zespół badawczy, bez potrzeby pracochłonnego „interakcyjnego” uzgadniania wykorzystanych kodów. Możliwe jest również zastosowanie takiej procedury agregacji kodów, która maksymalizowałaby rzetelność kodowania w sposób czysto matematyczny, bez potrzeby formułowania konkurencyjnych propozycji agregacji przez samych badaczy. Osobną kwestią stanowiłaby tu, oczywiście, łatwość czy nawet możliwość interpretacji tak wygenerowanych danych zagregowanych.

ków powinna być tak zbieżna, jak to tylko możliwe. W przypadku bardziej złożonych tekstów, twierdzenie to można jednak kwestionować, twierząc, że triangulacja powinna polegać właśnie na uwzględnieniu w badaniu *różnych* perspektyw na ten sam tekst.

Przypadek kodowania wypowiedzi Trybunału Konstytucyjnego jest tego dobrą ilustracją. Jak to zostało powiedziane, uzasadnienia orzeczeń TK charakteryzują się bardzo dużą złożonością i mogą podlegać różnorodnym interpretacjom, z których trudno jest wybrać interpretację ostateczną czy „prawidłową”. Istnieje ku temu kilka powodów. W tym samym fragmencie swojego uzasadnienia TK może wykorzystywać kilka typów argumentów czy metod wykładni prawa jednocześnie i może to robić z różną intensywnością. Niekiedy akcenty położone na poszczególne kwestie mogą być bardzo subtelne. Dlatego różni badacze, którzy mogą przeprowadzać kodowanie na różnym poziomie szczegółowości, mogą dostrzegać wszystkie lub tylko niektóre aspekty tego samego fragmentu tekstu, co będzie prowadzić do rozbieżności mierzalnych za pomocą miar rzetelności. Nawet w sytuacji, w której kodowanie tego samego fragmentu nie jest wielokrotne (tj. na jego opisanie użyto tylko jednego kodu, a nie kilku), te same wypowiedzi mogą też nie być konkluzywne i w konsekwencji mogą być kodowane za pomocą różnych kodów. Zatem w przypadku zrealizowania badania pozbawionego triangulacji, uzyskane dane będą silnie zniekształcone pod wpływem indywidualnej percepcji badacza. W konsekwencji, w procedurze badawczej wykorzystującej triangulację, wyższy poziom niezgodności pomiędzy badaczami może być przejawem nie tyle rzeczywistych odmienności pomiędzy nimi, co nieostrości wypowiedzi poddanych badaniu.

Warto również zauważyć, że uzasadnienia orzeczeń są wytwarzane we względnie złożonym procesie społecznym, co może negatywnie odbijać się na ich komunikatywności i ścisłości. Skład sędziowski rozstrzygający daną sprawę może, w przypadku TK, liczyć nawet kilkunastu orzeczników. Wprawdzie za przygotowanie pisemnego uzasadnienia wyroku jest formalnie odpowiedzialny jeden sędzia (w szczególnie złożonych sprawach może być ich więcej), ale w rzeczywistości w procesie tym udział bierze więcej osób. Wstępne wersje dokumentu są przygotowywane przez asystentów, a ostateczna treść uzasadnienia jest przedmiotem negocjacji (jeśli nie przetargu) pomiędzy poszczególnymi sędziami wchodzącymi w skład panelu orzekającego. Z tego względu ostateczny dokument może zawierać niespójności wewnętrzne, luki i różnego rodzaju niejasności.

Co być może jest jeszcze bardziej istotne, kody wykorzystane w opisywanych badaniach nie są zupełnie ostre. Jest to efekt wspomnianej wielości podejść teoretycznych, które zostały, *non lens volens*, zaimportowane do badania w efekcie zastosowania opisaną wyżej procedurę wytwarzania kodów przez badaczy będących z wykształcenia prawnikami i specjalizujących się w teorii prawa. Posługiwali się oni, w naturalny dla nich sposób, kategoriami znanymi tej dziedzinie wiedzy, a nie pojęciami całkowicie sztucznymi czy wytworzonymi w ramach teorii socjologicznej³¹. W konsekwencji, kody te są

³¹ Jest rzeczą wątpliwą, czy tego rodzaju kategorie teoretyczne, na obecnym etapie rozwoju socjologii prawa i socjologii teoretycznej, w ogóle istnieją. Warto tu odwołać się do obserwacji Bruno Latoura, który kwestionując możliwość ich istnienia, w następujący sposób scharakteryzował swoją pracę nad działanością francuskiej Conseil d'État: „[w]ydaje się, że w prawie nie ma żadnego stopniowania: albo się w nim jest bez reszty, albo wcale i mówi się o innych rzeczach. Jego warunki fortunnności (*conditions of felicity*)

niekiedy obciążone istotnym ładunkiem znaczeniowym.

Wreszcie, wiele z zastosowanych kategorii ma charakter raczej profesjonalnej wiedzy habitualnej, skodyfikowanej i zinstytucjonalizowanej w ramach profesji prawniczych, czy wręcz ideologicznej w sensie Mannheim'a, niż wiedzy ściśle teoretycznej, to znaczy dającej się sprowadzić do w pełni eksplikowalnych założeń epistemologicznych, ontologicznych czy deontologicznych. W efekcie niektóre przypadki niektórych typów wykładni mogą dać się zakwalifikować również jako przypadki innych typów wykładni. Przykład tego stanu rzeczy mogą stanowić niektóre typy tak zwanej wykładni językowej.

Istotę zagadnienia stanowi tu fakt, że w badaniach uzasadnień orzeczeń sądowych uniknięcie stosowania tego typu kategorii może nie być możliwe, a jeśli się to stanie, może prowadzić do negatywnych skutków poznawczych. Możliwe jest, rzecz jasna, zaprojektowanie takich narzędzi badawczych, które będą ściśle w tym sensie, że ich interpretacja *in abstracto* nie będzie rodziła sporów pomiędzy koderami. Można to uczynić choćby porzucając metodę kodowania *in vivo* i definiując kody w sposób arbitralny i dedukcyjny lub wykorzystując do ich wytwarzania osoby niemające wykształcenia prawniczego i niedysponujące odpowiednim kapitałem symbolicznym. Rodzi to jednak zagrożenie, że analiza orzecznictwa sądowego z wykorzystaniem takich kategorii mogłaby nie być w ogó-

lufortunności są wyjątkowo ostre. Etnograf zauważył to wielokrotnie, uświadamiając (...) sobie skalę własnej niemożności, tego by nawet po latach studiów, stopniowo zbliżyć się do wypowiedzi prawniczych. Mówienie językiem prawa było dla niego nieosiągalne nie tylko z braku słów i pojęć, ale z braku wszystkiego, absolutnie wszystkiego. Aby stwierdzić coś po prawniczymu, musiałby zostać radcą stanu. (...) Aby opisać Prawo w sposób przekonujący, trzeba do niego już wcześniej wskoczyć" (2010: 255 [tłum. własne]).

le możliwa lub nie doprowadziłaby do zrozumiałych wyników. W każdym razie, z definicji, opierałaby się na zerwaniu ze zbiorem kategorii znanych dyskursowi, do którego się odnosi, a byłaby dokonywana w pojęciach trudnych do translacji. Z tego względu jej wyniki mogłyby być niekomunikowalne w tym dyskursie.

Dlatego można byłoby sądzić, że uzyskanie wysokich miar rzetelności w badaniu na podobnym materiale zrealizowanym zgodnie z opisywaną procedurą badawczą wskazywałyby na zbliżone profesjonalne habitusy uczestników badania. W konsekwencji, osoba pochodząca spoza środowiska badaczy, kierując się inną ideologią czy zapatrywaniami teoretycznymi, mogłaby zinterpretować te same fragmenty orzeczeń zupełnie inaczej. Istnienie rozbieżności pomiędzy koderami może natomiast wskazywać na to, że wynik badania jest ostatecznie bardziej przekonujący, ponieważ koderzy odnosili się do szerszego zbioru przekonań i możliwości interpretacyjnych niż tylko jeden, partykularny punkt widzenia. Ten wniosek byłby zapewne tym bardziej przekonujący, im większa i bardziej zróżnicowana byłaby grupa badaczy³².

Najbardziej może interesujący wniosek, jaki płynie z tych obserwacji, dotyczy jednak ideologicznego charakteru kategorii pojęciowych wykorzystanych w omawianym badaniu. Zaobserwowane w badaniu istotne rozbieżności w kodowaniu tych samych fragmentów tekstu przez różnych badaczy ujawniają mianowicie to, że pozornie jednoznaczne kategorie stosowane

³² W tym kontekście, jako swoisty postulat metodologiczny, można byłoby sugerować swoistą kalibrację wykorzystywanych narzędzi poprzez syntetyczny pomiar różnic i podobieństw między habitusami badaczy w zakresie kodowanego materiału. W przypadku opisywanego tu schematu badawczego, wymagałoby to przeprowadzenia dodatkowej fazy badania: po ustaleniu listy kodów, należałoby przeprowadzić drugą fazę pilotażu, polegającą na kodowaniu dobranych celowo, tych samych orzeczeń TK.

przez prawników do opisywania podejmowanych w tym środowisku działań umysłowych w rzeczywistości są dość rozmyte i w znacznym zakresie podlegają interpretacji. Fakt ten pozostaje niezauważalny dopóty, dopóki nie przeprowadzi się systematycznego zestawienia użyć takich kategorii, dokonanych przez kompetentnych przedstawicieli tego zawodu.

Ta z kolei obserwacja daje się uogólnić poza problematykę socjologicznych badań nad orzecznictwem sądowym. Może mianowicie prowadzić do wątpliwości dotyczących wykorzystania wiedzy *insiderskiej* – czy szerzej, *insiderskiego* habitusu – w badaniach wykorzystujących metodologię postępowania badawczego, obcą temu habitusowi. Zmuszając koderów-*insiderów* do rezygnacji ze zwykłego sposobu użycia własnych kategorii i zastępując go systematyczną metodologią badań społecznych, niejako wyostrza się i unaocznia różnice pomiędzy rozumieniem tych samych pojęć przez różnych przedstawicieli tego samego środowiska. Innymi słowy, metodologia badań, z jej systematycznymi procedurami analizy treści, zwłaszcza wymuszonymi przez narzędzia informatyczne w CAQDA, występuje w roli analogicznej do Garfinkelowskiego wandalizmu interakcyjnego (Garfinkel 2007: 58 i nast.).

Taka alienacja badaczy od własnych, środowiskowych pojęć może prowadzić do ujawnienia ideologicznej natury tych kategorii, a efekt ten daje się mierzyć za pomocą współczynników rzetelności kodowania. Jest to rzecz jasna problem, który wymaga rozważenia za każdym razem, kiedy próbuje się zastosować *insiderskie* kategorie wypracowane w ramach grup zawodowych do prowadzenia *outsiderskich* badań praktyk dyskursywnych prowadzonych

w takich grupach. Obserwację tę można ująć w następujący trylemat. Po pierwsze, można sądzić, że zjawisko alienacji będzie tym bardziej widoczne, im bardziej rygorystyczne będą procedury analizowania i pomiaru rzetelności kodowania. Zastosowanie oprogramowania CAQDA niewątpliwie może się do tego przyczynić, unaoczniając różnice znaczeniowe zapoznawane w toku zwykłego dyskursu. Z drugiej strony, wypracowanie metodologii badania, która nie będzie prowadzić do alienacji, może negatywnie ciążyć na wiarygodności uzyskanych wyników. Wreszcie, po trzecie, zastosowanie kategorii zupełnie obcych badanemu dyskursowi może nie przynieść spodziewanych efektów poznawczych.

Wnioski

Zaprezentowane obserwacje i doświadczenia powinny przekonywać, że realizacja badań o opisanym, szerokim zakresie i przedstawionych założeniach poznawczych nie byłaby możliwa bez wykorzystania oprogramowania CAQDA. Narzędzia wykorzystane w badaniu pozwoliły zwłaszcza na jego przeprowadzenie w formule mieszanej, jakościowo-ilościowej analizy treści z kodowaniem swobodnym i triangulacją. Jakkolwiek możliwość kwantyfikacji wyników nie musi stanowić sama w sobie zalety, to realizacja badań w skali przedstawionej w artykule, i przy ograniczonych zasobach, jakie były do dyspozycji ich autorów, jest trudna do wykonania tylko w formie jakościowej. Na uwagę zasługuje tu zwłaszcza możliwość obliczania rzetelności kodowania przy zachowaniu swobodnego charakteru kodowania oraz duża elastyczność zastosowanego oprogramowania, gdy chodzi o sposób prezentacji wyników w formie syntetycznych wskaźników.

Nie ulega też wątpliwości, że opisywane badania nie wyczerpały całego zakresu form badania i możliwości analizy oferowanych przez zastosowane w nich oprogramowanie. Za opcje godne bliższego rozpoznania należy zwłaszcza uznać możliwość automatyzacji niektórych działań badawczych, prowadzenia badań w większych i rozproszonych zespołach, a także wykorzystania środowiska WWW do prezentacji wyników *in vivo*.

Negatywną stroną zestawu oprogramowania zastosowanego w badaniach jest konieczność posiadania umiejętności obsługi pakietu R, co wymaga znajomości podstawowych technik programistycznych. Wykorzystanie wszystkich możliwości oferowanych przez ten system, a zwłaszcza tworzenie procedur obliczeniowych, które byłyby efektywne pod względem wykorzystania zasobów komputerów, na jakich są uruchamiane (co może być istotne na przykład w przypadku dużych badań realizowanych przez rozproszony zespół badawczy albo środowiska WWW), wymaga natomiast pogłębionej wiedzy takiego rodzaju.

Bibliografia

Alexy Robert (1985) *Theorie der Grundrechte*. Baden-Baden: Nomos.

----- (1996) *Theorie der juristischen Argumentation: Die Theorie des rationalen Diskurses der juristischen Begriffs*. Frankfurt am Main: Suhrkamp.

Campbell John L. i in. (2013) *Coding In-Depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement*. „Sociological Methods and Research”, vol. 42(3), s. 294–320.

Stosunkowo duże wymagania pakietu R nie powinny jednak przysłonić faktu, że opisywany projekt dowiódł względnej łatwości korzystania z graficznej nakładki RQDA także przez użytkowników, którzy nie mają większego doświadczenia w pracy z oprogramowaniem tego rodzaju. Wreszcie, godną podkreślenia zaletą opisywanego oprogramowania jest jego dostępność na wolnej licencji i związana z tym nieodpłatność.

Prowadzone badania, oprócz otrzymania względnie nowatorskich wyników i uzyskania praktycznej wiedzy o stosunkowo mało znanym narzędziu badawczym, doprowadziły także do zadania interesujących pytań metodologicznych. Dotyczą one tak strony czysto technicznej badań – obliczania rzetelności kodowania w swobodnej analizie treści, jak i zagadnień epistemologicznych, związanych z możliwościami badania wyspecjalizowanego dyskursu profesjonalnego w kategoriach znanych temu dyskursowi i mu obcych. Obydwie te kwestie trudno uznać za ostatecznie rozwiązane w toku opisywanych badań, co dowodzi, że powinny być one przedmiotem dalszego namysłu.

Carey James W., Morgan Mark, Oxtoby Margaret J. (1996) *Intercoder Agreement in Analysis of Responses to Open-Ended Interview Questions: Examples from Tuberculosis Research*. „Cultural Anthropology Methods”, vol. 8(3), s. 1–5.

Carey James W. i in. (2004) *Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research*. „Field Methods”, vol. 16(3), s. 307–331.

Feteris Eveline T. (1999) *Fundamentals of Legal Argumentation: a Survey of Theories on the Justification of Judicial Decisions*. Hague: Kluwer.

Galligan Denis J., Matczak Marcin (2005) *Strategie orzekania sądowego o wykonywaniu władzy dyskrecyjnej przez sędziów sądów administracyjnych w sprawach gospodarczych i podatkowych*. Warszawa: Ernst & Young.

Garfinkel Harold (2007) *Studia z etnometodologii*. Przełożyła Alina Szulżycka. Warszawa: Wydawnictwo Naukowe PWN.

Huang Ronggui (2012) *RQDA: R-Based Qualitative Data Analysis* [dostęp 28 września 2013 r.]. Dostępny w Internecie <<http://rqda.r-project.org>>.

Kelsen Hans (2009) *Istota i rozwój sądownictwa konstytucyjnego*. Przełożył Bolesław Banaszkiewicz. Warszawa: Trybunał Konstytucyjny.

Krippendorff Klaus (2004a) *Content Analysis An Introduction to Its Methodology*. Thousand Oaks: Sage.

----- (2004b) *Reliability in Content Analysis: Some Common Misconceptions and Recommendations*. „Human Communication Research”, vol. 30(3), s. 411–433.

----- (2008) *Testing the Reliability of Content Analysis Data* [w:] Krippendorff Klaus, Bock Mary Angela, eds., *The Content Analysis Reader*. Thousand Oaks: Sage, s. 350–357.

Kurasaki Karen S. (2000) *Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data*. „Field Methods”, vol. 12, s. 179–194.

Latour Bruno (2010) *The Making of Law: An Ethnography of the Conseil d'Etat*. Cambridge: Polity Press.

Lautmann Rüdiger (1972) *Justiz – Die Stille Gewalt*. Frankfurt am Main: Suhrkamp.

Lombard Matthew, Snyder-Duch Jennifer, Campanella Bracken Cheryl (2002) *Content Analysis in Mass Communication Assessment and Reporting of Intercoder Reliability*. „Human Communication Research”, vol. 28(4), s. 587–604.

Maravall Jose Maria (2003) *The Rule of Law as a Political Weapon* [w:] Przeworski Adam, Maravall José María, eds., *Democracy and Rule of Law*. Oxford: Oxford University Press, s. 261–301.

Morawski Lech (2010) *Zasady wykładni prawa*. Toruń: TNOiK „Dom Organizatora”.

Neuendorf Kimberly A. (2002) *The Content Analysis Guidebook*. Thousand Oaks: Sage.

R Core Team (2012) *R: A Language and Environment for Statistical Computing* [dostęp 28 września 2013 r.]. Dostępny w Internecie <<http://www.R-project.org>>.

Ripley Brian D. (2001) *The R Project in Statistical Computing*, „MSOR Connections. The Newsletter of the LTSN Maths, Stats & OR Network”, vol. 1(1), s. 23–25.

Sadurski Wojciech (2008) *Rights before Courts: A Study of Constitutional Courts in Postcommunist States of Central and Eastern Europe*. Dordrecht: Springer.

Scheffer Thomas (2010) *Adversarial Case Making. An Ethnography of English Crown Court Procedure*. Leiden: Brill.

Stawecki Tomasz, Winczorek Jan, red., (w druku) *Wzorce wykładni konstytucji w Polsce i państwach Europy Środkowej – doktryna i praktyka*. Warszawa: Wolters Kluwer.

Stone Sweet Alec (1999) *Judicialization and the Construction of Governance*. „Comparative Political Studies”, vol. 32(2), s. 147–184.

Winczorek Jan, Stawecki Tomasz, Staśkiewicz Wiesław (2008) *Between Polycentrism and Fragmentation: The Impact of Constitutional Tribunal Rulings on the Polish Legal Order*. Warszawa: Ernst & Young.

Wronkowska Sławomira (2008) *Kilka uwag o „prawodawcy negatywnym”*. „Państwo i Prawo”, t. 64(10), s. 5–21.

Wróblewski Jerzy (1988) *Sądowe stosowanie prawa*. Warszawa: PWN.

Wyrembak Jarosław (2009) *Zasadnicza wykładnia znamion przestępstw: pozycja metody językowej oraz rezultatów jej użycia*. Warszawa: Wolters Kluwer.

Zieliński Maciej (2008) *Wykładnia prawa: zasady, reguły, wskazówki*. Warszawa: LexisNexis.

Aneks

Tabela 1. Lista kodów wraz ze schematem ich agregacji. Kursywą zaznaczono kody drugiego poziomu.

KOD PO AGREGACJI	KODY PRZED AGREGACJĄ
argument językowy	argumenty semantyczne – argument semantyczny inny argumenty semantyczne – argument z autonomicznego znaczenia pojęć argumenty semantyczne – argument z języka naturalnego argumenty semantyczne – argument z języka prawnego argumenty semantyczne – argument z niejasności przepisu przywołanie tekstu normatywnego
argument systemowy	argumenty systemowe – argument z metod regulacji i gałęzi prawa argumenty systemowe – argument z typu przepisu argumenty systemowe – argument z założeń idealizacyjnych systemu prawa argumenty systemowe – argument systemowy inny argumenty systemowe – argumenty ze struktury systemu prawa argument z faktu trwałości lub zmiany prawa argument z formalnego obowiązywania prawa klasyczne argumenta prawnicze
argument systematyczny	argumenty systemowe – argumenty z budowy aktu prawnego
argument funkcjonalny	argument z faktów powszechnie znanych argument z istoty instytucji prawnej argument ze skutków społecznych argument ze zmiany społecznej
argument celowościowy i intencjonalny	argument z ratio legis odwołanie do prac legislacyjnych
argument komparatystyczny	argumenty komparatystyczne – argument komparatystyczny wewnętrzny argumenty komparatystyczne – argument komparatystyczny zewnętrzny
argument z własnego autorytetu	argumenty z orzecznictwa – argument z własnego orzecznictwa TK – bez uzasadnienia argumenty z orzecznictwa – argument z własnego orzecznictwa TK z uzasadnieniem <i>krytyka lub odrzucenie argumentu</i>
argument z autorytetu innego sądu	argumenty z orzecznictwa – argument z orzecznictwa sądów polskich argumenty z orzecznictwa – argument z orzecznictwa ETPCz argumenty z orzecznictwa – argument z orzecznictwa ETS argumenty z orzecznictwa – argument z orzecznictwa innego sądu zagranicznego
argument z doktryny	argument z doktryny

argument z dystrybucji władzy	argument z dyskrecjonalności organów władzy argument z zakresu władztwa TK
argument z interesu (publicznego, grupowego, indywidualnego)	odwołanie do interesów grupowych lub grup społecznych
argument z zasad prawa lub wartości prawnych i pozaprawnych	argument z zasady interpretowanej argument z zasady rozstrzygającej interpretację argument z proporcjonalności i wagi (zasad, wartości) odwołanie do wartości i innych norm pozaprawnych
<i>nie-argumenty (fragmenty uzasadnień wyłączone z analizy)</i>	<i>argumentum ad rem</i> <i>elementy orzeczenia – konkluzja rozumowania</i> <i>elementy orzeczenia – powtórzenie</i> <i>elementy orzeczenia – zdanie odrębne</i> <i>narzędzia retoryczne</i>
<i>konstytucja</i>	<i>typy aktów prawnych – konstytucja</i>
<i>krajowe akty normatywne inne niż konstytucja</i>	<i>typy aktów prawnych – non-state law</i> <i>typy aktów prawnych – prawo europejskie pierwotne</i> <i>typy aktów prawnych – prawo europejskie wtórne</i> <i>typy aktów prawnych – prawo międzynarodowe publiczne</i> <i>typy aktów prawnych – ustawa</i> <i>typy aktów prawnych – akty nieustawowe</i>

Źródło: opracowanie własne.

Cytowanie

Winczorek Jan (2014) *Wykorzystanie oprogramowania R i RQDA w jakościowo-ilościowej analizie treści orzeczeń Trybunału Konstytucyjnego*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 138–161 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

Usage of Software Packages R and RQDA in a Qualitative-Quantitative Content Analysis of Verdicts of Polish Constitutional Tribunal

Abstract: The paper describes utilization of R statistical software with RQDA CAQDA package in a mixed qualitative-quantitative triangulated content analysis with free coding, intercoder reliability analysis, and statistical significance analysis, performed on a sample of justifications of verdicts of Polish Constitutional Tribunal, issued between 1986 and 2009. It presents the most important features of software used, assumptions of the study, its methodology and procedures of data analyses, as well as epistemic questions which arose in the course of its execution.

Keywords: CAQDA, RQDA, R (software), content analysis, intercoder reliability, Polish Constitutional Tribunal

Kamil Głowacki
Uniwersytet Łódzki

DOI: <https://doi.org/10.18778/1733-8069.10.2.09>

Organizacja procesu badawczego a oprogramowanie do organizacji wiedzy i zarządzania projektem badawczym

Abstrakt Artykuł wprowadza czytelnika w obszar programów komputerowych, które mogą wspomagać pracę nad projektem badawczym w poszczególnych jego fazach zarówno na poziomie organizacyjnym, jak i merytorycznym. Celem artykułu nie jest porównanie i przedstawienie wszystkich programów, a jedynie zaprezentowanie pojedynczych przedstawicieli grup programów, które można zestawić w jeden działający system zarządzania projektem, organizacji wiedzy oraz danych bibliograficznych, a także analizy materiału empirycznego z wykorzystaniem rodziny programów CAQDAS (Computer Assisted Qualitative Data Analysis Software).

Słowa kluczowe CAQDAS, organizacja i zarządzanie wiedzą, zarządzanie danymi bibliograficznymi, oprogramowanie wspomagające proces badawczy, metoda teorii ugruntowanej

Kamil Głowacki, doktorant w Katedrze Socjologii Organizacji i Zarządzania w Instytucie Socjologii Uniwersytetu Łódzkiego. W obszarze jego zainteresowań badawczych znajdują się: symboliczny interakcjonizm, metodologia teorii ugruntowanej, socjologia wizualna i jej wykorzystanie do badania procesów społecznych, fotografia jako działanie społeczne, a także zastosowanie nowych technologii w badaniach jakościowych.

Adres kontaktowy:

Katedra Socjologii Organizacji i Zarządzania
Instytut Socjologii, Wydział Ekonomiczno-Socjologiczny
Uniwersytet Łódzki
ul. Rewolucji 1905 r. 41, 90-214 Łódź
e-mail: głowacki.kamil@gmail.com

Współczesny rozwój technologii komputerowej, generowanie wiedzy w postaci cyfrowych publikacji, digitalizacja ich papierowych wersji oraz dostępność coraz bardziej zróżnicowanych źródeł danych skłaniają do refleksji nad wykorzystywaniem w warsztacie badacza nauk społecznych nowych możliwości gromadzenia wiedzy i materiałów referencyjnych, jak i organizacją samego procesu badawczego¹. Jednak sposób, w jaki każdy badacz organizuje swój styl pracy nad projektem badawczym, w dużej mierze zależy od jego własnych preferencji, wybranego obszaru badań, wybranej metodologii czy postawio-

¹ Na wstępie należy zastanowić się, co rozumiemy pod pojęciem procesu badawczego. Na potrzeby niniejszego artykułu jest to zespół czynności i decyzji, jakie musi podjąć badacz w trakcie pracy.

negu problemu badawczego. Co więcej – praca nad projektem badawczym to nie tylko praca organizacyjna, ale również koncepcyjna. Obie te sfery przenikają się nawzajem i współwystępują na każdym etapie badania, dlatego w poniższym artykule przedstawię narzędzia wspierające oba te procesy. Na potrzeby artykułu przyjmujemy, iż rozważamy realizację projektu opierającego się na metodach jakościowych z wykorzystaniem metody teorii ugruntowanej. Nie oznacza to jednak, że proponowanych narzędzi nie można użyć przy innych projektach badawczych.

Pierwszym etapem każdego projektu jest praca merytoryczna nad wyborem obszaru, który będzie podlegał dalszej eksploracji, oraz analiza ram formalnych, w jakich projekt może być realizowany. W ramach tego etapu następuje również ustalenie struktury organizacyjnej projektu, czyli: negocjowany jest podział pracy, formułowany wstępny harmonogram oraz szacowane są koszty realizacji projektu. Równolegle prowadzone są prace koncepcyjne nad problemowym zakresem badania.

Drugi etap realizacji projektu rozpoczynamy od przeglądu dostępnego dorobku naukowego, najczęściej poprzez zapoznanie się z dostępną literaturą przedmiotu. Ten proces ma miejsce nawet jeśli korzystamy z metodologii teorii ugruntowanej, bo choć zgodnie z założeniami Glasera i Straussa (1999) badacz nie powinien przystępować do badania ze wstępnymi prekonceptualizacjami, to jednak nie może on pozostać ignorantem wobec dorobku społeczności naukowej. Uzasadnienie takiego postępowania odnajdujemy u Kuhna (2001), który rekonstruując strukturę powstawania teorii naukowych, pokazuje jak ważne jest poznanie dorobku społeczności naukowej, w ramach której podejmuje się pracę badawczą.

Etap trzeci to przystąpienie do zbierania danych i ich analiza. Stosując MTU, te dwa działania są prowadzone jako równoległe strumienie wzajemnie na siebie oddziałujące. Jednocześnie prowadzą one do próbkowania teoretycznego i wysycenia zbieranego materiału. W tym samym czasie analityk przystępuje najpierw do kodowania otwartego, aby w dalszym kroku rozpocząć scalanie kategorii i tworzenie kodów teoretycznych. Proces ten jest prowadzony naprzemiennie z tworzeniem notatek wyjaśniających decyzje kodowe oraz not teoretycznych wskazujących na związki pomiędzy poszczególnymi kodami i kategoriami. Wszystkie te elementy wraz z zastosowaniem metody porównywania mają doprowadzić badacza do stworzenia teorii średniego zasięgu, która będzie teorią pracującą i wyjaśniającą.

Etap czwarty to opracowanie wyników w formie przystępnej do dalszej obróbki i publikacji. Przyjmuje on często formę tworzenia raportów, artykułów oraz książek zawierających wyniki naszych badań. Jest to etap związany z upowszechnianiem wyników opracowanych w ramach realizowanego projektu.

Ten kilkietapowy proces może być dość wymagający, szczególnie jeśli badanie prowadzone jest w ze-spole badaczy rozrzuconych po całym kraju czy też pracujących w różnych ośrodkach zagranicznych. Niemniej jednak można go usprawnić, przyspieszyć i uporządkować za pomocą konkretnych narzędzi komputerowych, co jednocześnie pozwoli na jego większą kontrolę. W niniejszym artykule przedstawię kilka z nich, wskazując jednocześnie, w którym etapie procesu badawczego należy je zastosować. W prostym schemacie obrazuje to tabela 1.

Tabela 1. Etapy projektu badawczego wraz z dopasowanym oprogramowaniem wspierającym.

ETAPY REALIZACJI PROJEKTU BADAWCZEGO	PROPONOWANE OPROGRAMOWANIE
Etap 1 – merytoryczno-organizacyjny	Mind42 Freemind ProjectLibre
Etap 2 – praca z literaturą	Citavi Mendeley
Etap 3 – zbieranie i analiza danych jakościowych	NCapture Nvivo
Etap 4 – upowszechnianie wyników w formie publikacji	Scrivener Wybrany edytor tekstu – MS Word

Źródło: opracowanie własne.

Należy w tym miejscu jasno zaznaczyć, że prezentowane narzędzia stanowią autorską propozycję i zostały wybrane w sposób subiektywny, w związku z czym nie powinny być traktowane jako zamknięty katalog, a jedynie jako impuls do własnych poszukiwań i testów. Rynek obfituje bowiem w liczne rozwiązania programowe, których porównań nie sposób przytoczyć w niniejszym artykule.

Etap 1 – praca merytoryczno-organizacyjna

Praca koncepcyjna – mapy myśli

Rozpoczynając pracę koncepcyjną, możemy wykorzystać techniki wyzwalające kreatywność i pozwalające na wyjście poza znane schematy poznawcze. Jedną z takich technik jest burza mózgów. Jedno-

nocześnie należy pamiętać, iż technika ta to nie tylko generowanie pomysłów zapisywanych na tablicy, ale także dość mocno ustrukturalizowana technika wyzwalania potencjału grupy. Nie jest ona jednak wolna od błędów, które mogą zakłócić proces generowania rozwiązań i zbierania najlepszych pomysłów. Kluczowym zagrożeniem, jakie może wkraść się w trakcie pracy z burzą mózgów, jest błąd myślenia grupowego i wpływ autorytetów na wyniki (Aronson 2001).

Narzędziami, jakie mogą zastąpić nam tablicę i jednocześnie nadawać strukturę prowadzonej pracy koncepcyjnej, są programy do mapowania myśli. Jedną z interesujących propozycji jest MindMap opracowywany przy współpracy z twórcą map myśli Tonyem Buzanem. Jest to propozycja płatna, o dość rozbudowanym interfejsie. Rozwiązaniem zdecydowanie prostszym i dostępnym jako bez-

płatna aplikacja webowa jest Mind42. Jest to prosty w obsłudze program wymagający jedynie założenia konta na serwerze produktu. Główną zaletą tego oprogramowania jest jego prostota oraz możliwość uruchomienia na każdym urządzeniu podłączonym do Internetu. Taka konstrukcja Mind42 pozwala na gromadzenie map myśli na serwerach wydawcy programu, a tym samym ich współtworzenie, współdzielenie oraz udostępnianie, jak i publikowanie w Internecie. Współtworzenie jest możliwe w czasie rzeczywistym, co oznacza, że w jednej chwili na otwartym dokumencie może pracować więcej niż jeden użytkownik. Pozwala to na prowadzenie burzy mózgów czy pracy koncepcyjnej wielu osób bez konieczności ich fizycznej obecności w jednej przestrzeni. Olbrzymią zaletą opisywanego programu jest możliwość zapisywania opracowywanej mapy w postaci plików o formacie *.mm, które to rozszerzenie obsługiwane jest przez programy takie jak FreeMind, Docear, Citavi, Scrivener i tym podobne. Oznacza to, iż praca wykonana w programie Mind42 może być użyta w bardziej rozbudowanych programach do pracy koncepcyjnej, generowania i pokazywania związków między pojęciami, ale może także posłużyć do utworzenia drzewa kategorii w programach do organizowania wiedzy i zarządzania bibliografią, jak również w studiach pisarskich do tworzenia konspektu raportu czy dysertacji. Ta funkcjonalność pozwala traktować Mind42 jako narzędzie, które nie tylko w łatwy sposób pozwala zapisywać pojawiające się pomysły i koncepcje w postaci cyfrowej, ale także jako narzędzie stanowiące część warsztatu badacza, w ramach którego stworzony materiał może być dowolnie rozwijany w dalszych fazach pracy badawczej.

Inną grupą programów do tworzenia map myśli są programy pozwalające na tworzenie tak zwanych map konceptualnych. Są to mapy kategorii, które mogą pojawiać się w sposób niezależny od siebie i początkowo pozbawiony relacji. W miarę rozwijania projektu badawczego, badacz może dodawać związki pomiędzy pojęciami i systematycznie tworzyć coraz bardziej złożone mapy pojęć i relacji. Darmowym i typowym przykładem takiego oprogramowania jest CmapTools. Szeroki opis tego programu znajdziemy w publikacji Jakuba Niedbalskiego (2013).

Nieco innym oprogramowaniem, które także jest darmowe i pozwala tworzyć kategorie nieprzyporządkowane na początku procesu analizy, jest Docear. Jest to open source'owy projekt rozwijany przez niemieckich programistów, będący jedną z wersji innego darmowego programu Freemind, na bazie którego powstał. Docear to program dobrze współpracujący z menedżerami bibliografii, którego największą zaletą jest odczytywanie notatek tworzonych na plikach PDF, na przykład w programie Adobe Reader. Daje on także możliwości gromadzenia i organizacji wiedzy w postaci map myśli, których węzły mogą być linkowane do zewnętrznych dokumentów, takich jak artykuły czy publikacje w formacie PDF, strony Internetowe, pliki o rozszerzeniu .doc. W programie tym możemy także tworzyć notatki powiązane i mapy myśli z węzłami odwołującymi się do siebie nawzajem.

Cykl projektu badawczego

Na pierwszym etapie pracy z nowym projektem badawczym, kiedy już zdefiniowany zostanie

obszar badań i cel jego prowadzenia, warto zastanowić się nad strukturalno-organizacyjnym aspektem jego realizacji. W projektach naukowych realizowanych w ramach firm badawczych nie ma możliwości realizacji projektu bez uwzględnienia potrzebnych na jego realizację środków finansowych i oszacowania zaangażowania kapitału ludzkiego oraz zaplanowania systemu sprawozdawczości z realizowanego projektu. Takie podejście wymusza w fazie wstępnej analizę strukturalno-organizacyjną oraz ustalenie sposobów monitorowania postępu realizacji projektu. Wykorzystanie takiego podejścia do projektów badawczych realizowanych w ramach uczelni wyższych, niekoniecznie na zlecenie organizacji zewnętrznych, ale w celu powiększenia zasobu wiedzy o otaczającej nas rzeczywistości, może przyczynić się do zwiększenia efektywności całego procesu, ale także uzyskania zewnętrznego wsparcia naszych inicjatyw badawczych. Jeśli stoimy przed mniejszym projektem, jakim jest praca magisterska czy rozprawa doktorska, wykorzystanie dobrych praktyk związanych z zarządzaniem dużymi projektami badawczymi także pozwoli oszczędzić narastającego stresu w miarę zbliżania się wyznaczonego terminu zakończenia naszych przedsięwzięć. I o ile będąc socjologami, antropologami czy etnografami, odebraliśmy odpowiednie wykształcenie dotyczące wyboru właściwej metodologii naszego badania i etapów jego realizacji, o tyle w obszarze zarządzania tymi działaniami zwykle mamy lukę kompetencyjną. W tej części artykułu chciałbym więc podać kilka wskazówek zaczerpniętych z różnych metodyk zarządzania projektami² wraz

z podpowiedzią, jakich narzędzi możemy użyć, aby cały proces był dla nas jak najbardziej efektywny.

Skupiając się jedynie na najważniejszych elementach, możemy rozpocząć od stworzenia PID (Project Initiation Document). Jest to podstawowy dokument, który zawiera skrótowy opis naszego projektu. PID jest niezwykle przydatny przy podejmowaniu wstępnych rozmów z osobami, od których zależy dofinansowanie projektu i jego przebieg. W dokumencie tym zawieramy odpowiedzi na takie pytania, jak: Jakie są cele naszego projektu? Dlaczego należy je osiągnąć? Kto będzie odpowiedzialny za projekt? Kiedy i jak będzie prowadzony projekt?

Z kolei pierwszym narzędziem, od którego dobrze jest zacząć dalszą pracę nad projektem, jest strukturalny podział pracy (SPP [ang. Work Breakdown Structure]). Zastosowanie tego narzędzia pozwala na bardzo dokładne rozpisanie czynności do wykonania w projekcie, czego efektem jest realizacja założonych celów. SPP pozwala także na lepszą kontrolę nad procesem realizacji projektu. Korzystając z SPP, powinniśmy dążyć do rozłożenia naszych zadań na podstawowe elementy składowe. Innymi słowy, należy dążyć do rozbicia naszych głównych działań na pojedyncze czynności, które będą wykonywane przez jedną osobę w określonym czasie (*Practice Standard for Work Breakdown Structures* 2006; Głowacki 2011; Schwalbe 2011). Robiąc to, powinniśmy się skupić na rozkładzie czynności w czasie, nie zaś na przypisywaniu ich konkretnym członkom naszego zespołu. Rozpatrywanie zależności kilku czyn-

ności w czasie oraz odpowiedzialność za ich wykonanie będą analizowane dopiero w kolejnym kroku. W tym celu możemy wykorzystać matrycę odpowiedzialności (Responsibility Assignment Matrix), którą opracowujemy, korzystając ze stworzonego wcześniej SPP i opisu ról pełnionych w zespole projektowym. W wierszach matrycy umieszczamy czynności zgodnie z ich podziałem stworzonym w SPP, w kolumnach umieszczamy kolejnych członków zespołu. Następnie, zgodnie z zasadą RACI, rozdzielamy odpowiedzialność w obrębie czynności prowadzących do ostatecznego zrealizowania założonego celu. Skrót RACI pochodzi od pierwszych liter anglojęzycznych słów opisujących poziom odpowiedzialności (Głowacki 2011; *A guide to the Project Management Body of Knowledge* 2013):

R – Responsible – jest to osoba, która jest odpowiedzialna za osobiste wykonanie czynności;

A – Accountable – jest to osoba, która zatwierdza wykonanie czynności przez osobę oznaczoną jako Responsible;

C – Consult – jest to osoba, której opinia powinna zostać wzięta pod uwagę przy podejmowaniu decyzji lub wykonywaniu danej czynności;

I – Inform – jest to osoba, która powinna zostać poinformowana o wykonaniu danej czynności.

Dokument, który otrzymujemy po zakończeniu pracy z matrycą, pozwala nam ocenić dokładnie, jak będzie wyglądało zaangażowanie poszczególnych członków zespołu w realizację projektu. Poza tym

pomaga – w trakcie jego trwania – w kontrolowaniu wywiązywania się poszczególnych osób ze swoich zadań.

Oprogramowań wspierających wymienione powyżej narzędzia projektowe jest obecnie na rynku bardzo wiele. Od drogich i bardzo popularnych, takich jak MS Project, po darmowe i choć może nie tak estetyczne, to nadal funkcjonalne, takie jak na przykład ProjectLibre.

Ta ostatnia to desktopowa aplikacja, rozwijana od 2008 roku, pozwalająca na zarządzanie niedużym projektem. W obecnej fazie rozwoju aplikacja nie wspiera rozwiązań wieloprojektowych oraz pracy w chmurze, dlatego jej zastosowanie do pracy zespołowej jest ograniczone i możliwe tylko poprzez serwisy udostępniające dzielenie plików, jak na przykład Dropbox czy Storino (korzystając z tej opcji, musimy pamiętać, aby nie pracować nad plikiem równoległe z innym członkiem zespołu). Program możemy pobrać ze strony www.projectlibre.com. Znajdziemy tam także informacje, iż planowane jest udostępnienie wersji „chmurowej” oprogramowania w 2014 roku. Zatem, być może w chwili publikacji artykułu będzie można już z niej korzystać.

Praca z ProjectLibre jest dość intuicyjna. Interfejs główny jest ascetyczny i wykorzystuje wstążkę znaną z takich produktów, jak chociażby MS Word. W widoku domyślnym program otwiera wykres Gantta, w ramach którego rozpoczynamy tworzenie SPP z przypisaniem czasu, a w dalszej kolejności zasobów ludzkich i materialnych potrzebnych do zrealizowania projektu.

² Najbardziej rozpowszechnione metodyki to PMI, PCM, PRINCE2, AGILE.

Rys. 1. Wykres Gantta z podziałem projektu na zadania w programie ProjectLibre.

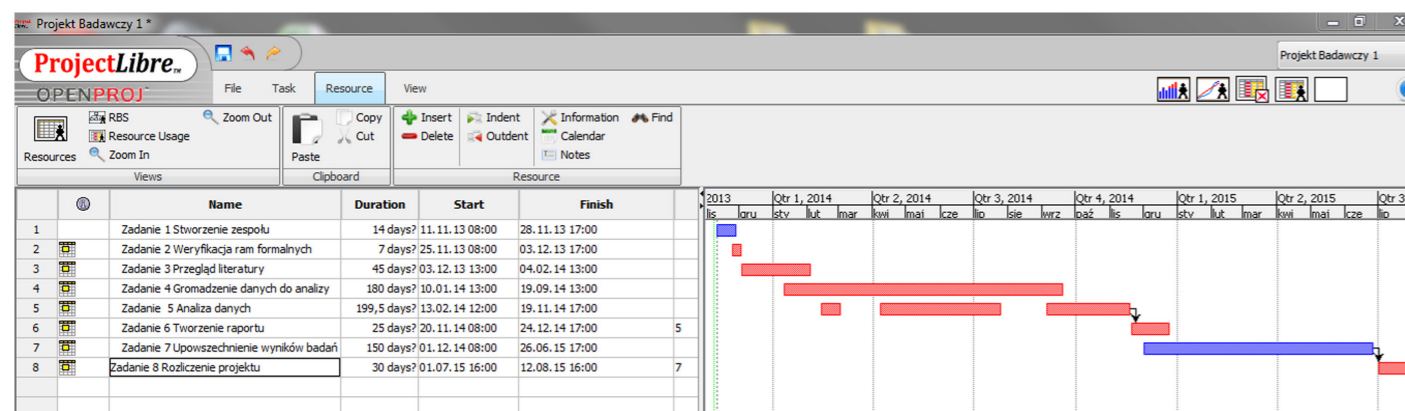


Źródło: opracowanie własne. Program ProjectLibre.

Powyższy zrzut ekranu pokazuje obszar roboczy, jaki mamy do dyspozycji. Do programu zostały wprowadzone podstawowe zadania do kolejnych etapów projektu badawczego. Do każdego zadania została przypisana liczba dni potrzebnych na jego wykonanie. Wykres po prawej stronie pokazuje, że

część zadań może być wykonywana równocześnie, a część jest od siebie zależna. Zgodnie z opisaną zasadą tworzenia strukturalnego podziału pracy, każde z zadań powinno zostać teraz podzielone na czynności, którym przypiszemy czas realizacji i osoby odpowiedzialne za ich wykonanie.

Rys. 2. Widok zadania podzielonego na czynności w ProjectLibre.



Źródło: opracowanie własne. Program ProjectLibre.

ProjectLibre pozwala nam również na śledzenie postępów w realizacji projektu oraz czasu, jaki musimy poświęcić w danym miesiącu, aby zrealizować zaplanowane czynności. Jest to bardzo przydatna funkcja, która na początku planowania projektu

pozwala nam dość dokładnie przyjrzeć się naszym założeniom i zweryfikować je. Przykładowo, jeśli w jednym czasie zaplanujemy przeprowadzenie wywiadów, zrobienie transkrypcji i obserwację, to program wskaże nam, że w danym miesiącu

będziemy potrzebowali na przykład 352 godzin roboczych na wykonanie zaplanowanych zadań. Wówczas musimy się zastanowić, czy taką pracę jest w stanie wykonać jedna osoba, czy musimy

przypisać do tych zadań dodatkowego badacza lub rozłożyć w czasie realizację tych czynności, co może skutkować wydłużeniem terminu realizacji naszego projektu poza wyznaczone granice.

Rys. 3. Rozkład czasu potrzebnego na wykonanie zadań w danych miesiącach – ProjectLibre.

Name	Work	Duration	Start	Finish
Zadanie 1 Stworzenie zespołu	112 hours	14 days?	11.11.13 08:00	28.11.13 17:00
Zadanie 2 Weryfikacja ram formalnych	56 hours	7 days?	25.11.13 08:00	03.12.13 17:00
Zadanie 3 Przegląd literatury	360 hours	45 days?	03.12.13 13:00	04.02.14 13:00
Zadanie 4 Gromadzenie danych do analizy	2 004 hours	143 days?	10.01.14 13:00	30.07.14 13:00
Z4C1 Przeprowadzenie 100 wywiadów swk	800 hours	100 days?	10.01.14 13:00	30.05.14 13:00
Z4C2 Transkrypcja wywiadów	228 hours	28,5 days?	31.05.14 08:00	10.07.14 13:00
Z4C3 Import do Nvivo	8 hours	1 day?	11.07.14 13:00	14.07.14 13:00
Z4C4 Przeprowadzenie obserwacji uczestniczących	960 hours	120 days?	10.02.14 13:00	28.07.14 13:00

Źródło: opracowanie własne. Program ProjectLibre.

Zachęcam wszystkich badaczy, którzy planują rozpoczęcie projektu badawczego, zarówno małego, jakim może być praca magisterska, jak i dużego, jak projekt realizowany przez uczelnię, do zapoznania się z możliwościami programów wspierających planowanie i realizację projektów, gdyż metodyczne podejście do planowanych zadań, ich nadzorowanie, rozliczanie i zakończenie w wyznaczonym czasie jest tak samo istotne jak poprawność metodologiczna. Szczególnie istotne jest to wówczas, kiedy są one finansowane ze źródeł zewnętrznych, a niedoszacowanie czasu i potrzebnych zasobów może wiązać się z konsekwencjami finansowymi.

Etap 2 – praca z literaturą

Powołania i bibliografia

Kolejnym etapem pracy badacza po wstępnym zarysowaniu problematyki badawczej jest przystąpienie do systematycznej analizy dostępnej lite-

ratury w obszarze naszych zainteresowań badawczych. Nie warto bowiem odpowiadać na pytania badawcze, na które ktoś już udzielił odpowiedzi, a budowanie nowej wiedzy wymaga od badacza zapoznania się z dorobkiem społeczności naukowej tworzącej w ramach dziedziny, którą uprawia. Jedną z podstawowych metod organizowania i pozyskiwania nowej literatury z danego obszaru są dane bibliograficzne. Tworząc teksty naukowe, autor jest zobowiązany do wskazania literatury, z której korzystał i która była dla niego inspiracją.

Powołanie na literaturę może przyjmować formę bezpośredniego cytowania, zapożyczenia materiału z przedstawieniem go w zmienionej postaci, która pasuje do reszty artykułu lub jedynie wzmianki zalecającej czytelnikowi sięgnięcie do dodatkowego, pełniejszego źródła informacji. W każdym przypadku autor jest zobowiązany do podania opisu bibliograficznego pracy, na którą się powołuje, przy czym odpowiednie

informacje nie tylko muszą wystarczać do odszukania źródła, ale także spełniać formalne wymagania wydawcy odnośnie ich zakresu, sposobu połączenia z tekstem oraz stylu prezentacji. (Parfieniuk 2012: 196)

Mimo wiekowej tradycji bibliograficznej i ogromnego jej rozwoju w XX wieku i na początku XXI wieku, nadal pozostaje w tym zakresie bardzo wiele do zrobienia pod kątem wdrażania wspólnych norm, stylów bibliograficznych i cytowań. W obecnej chwili funkcjonuje kilka norm, które mają systematyzować sposób tworzenia bibliografii i formę cytowania prac. Należą do nich na przykład norma PN-ISO 690:2012P zatwierdzona przez Polski Komitet Normalizacyjny. Jest ona polskim odpowiednikiem międzynarodowej normy ISO 690:2010 i zastępuje trzy wcześniej funkcjonujące normy. Jednak nadal wiele wydawnictw posiada swoje własne style prezentacji cytowań i przypisów. Aby uświadomić sobie olbrzymią ich różnorodność, wystarczy nadmienić, że jedno z oprogramowań do zarządzania bibliografią posiada w swojej bazie ponad 5 tysięcy stylów, a to nadal nie wyczerpuje wszystkich możliwości. Dlatego wszystkie kwestie formalno-organizacyjne związane ze złożeniem i korektą artykułu lub późniejszym wykorzystaniem części bibliografii w kolejnej publikacji powodują, że udział czasu poświęconego na jej edycję i poprawki staje się coraz większy. Do tego dochodzi jeszcze praca z rosnącą bazą danych bibliograficznych i cytowań, ich oznaczanie, porządkowanie i przeszukiwanie w celu dalszego wykorzystania, a niejednokrotnie także mozolna praca nad weryfikacją poprawności danych, które badacz wprowadzał ręcznie.

W niniejszym artykule chcę zwrócić uwagę czytelnika na narzędzia, które być może warto umieścić w swoim warsztacie, aby zmniejszyć ilość czasu, jaką badacz-autor musi poświęcić na poszukiwanie, gromadzenie nowej literatury i – w późniejszym etapie – dostosowywanie stylów cytowań i bibliografii do poszczególnych wydawnictw i czasopism. Tym sposobem otwiera się przed nami bardzo szerokie zagadnienie, jakim w pierwszej kolejności są bazy danych bibliograficznych i ich organizacja, a w dalszej części – cytowania i organizacja wiedzy.

Bazy danych bibliograficznych

Bazy danych bibliograficznych w ujęciu, jakie nas interesuje, to zdigitalizowane katalogi biblioteczne pozwalające na przeszukiwanie ich zasobów przy pomocy komputera, z możliwością segregowania danych po takich atrybutach, jak: rodzaj poszukiwanej publikacji, autor, tytuł, hasła przedmiotowe, rok wydania, miejsce wydania, aby wymienić te podstawowe.

Początki digitalizacji katalogów bibliotecznych w Polsce sięgają drugiej połowy lat sześćdziesiątych, kiedy w Bibliotece Narodowej próbowano wdrażać system wspomagający prace katalogowe i redakcyjne ARKA (Kolasa 2000). W tym okresie także Politechnika Wrocławska pracowała nad systemami z wykorzystaniem maszyn Odra 1300 i RIAD. W kolejnych latach do bibliotek uczelni wyższych systematycznie wkraczały systemy bazodanowe. Prace nad ich ulepszaniem i usprawnianiem nabrały rozmachu w latach dziewięćdziesiątych, co pokryło się ze wzrostem dostępności Internetu i możliwości łączenia poszczególnych

baz danych w ogólnodostępne katalogi, do których obecnie mamy dostęp za pośrednictwem sieci internetowej. Jednym z najbardziej popularnych formatów zapisu danych bibliograficznych są pliki tekstowe o rozszerzeniu .bib. Zawierają one rekordy różnych typów, na przykład: artykuł, czasopismo, książka, dokument internetowy, dokument niepublikowany. Treść rekordu składa się zawsze z klucza, który jest czymś w rodzaju nazwy skróconej, jaka pozwoli zidentyfikować właściwy rekord oraz pól o różnych nazwach i wartościach uzależnionych od typu dokumentu. Zastosowanie tego systemu plików pozwala na tworzenie globalnych katalogów, takich jak na przykład WorldCat, który jest systemem baz danych rozproszonych składającym się z ponad 71 000 katalogów bibliotek na całym świecie. Innym rodzajem baz danych jest na przykład Usługa EBSCOhost. Jest to usługa płatna, jednak wykupienie odpowiedniej subskrypcji daje możliwość dostępu do pełnych tekstów publikacji. EBSCOhost składa się z wielu tematycznych baz danych zawierających dane bibliograficzne z konkretnych obszarów, jak na przykład SocINDEX. Inną usługą oferującą dostęp do danych bibliograficznych i abstraktów jest Web of Science, będąca zbiorem kilku dużych baz danych, między innymi: Art&Humanities Citation Index, Conference Proceedings Citation Index-Science, Conference Proceedings Citation Index – Social Science and Humanities, Social Science Citation Index. Usługa ta jest oferowana przez Web of Knowledge i opłacana przez Ministerstwo Nauki i Szkolnictwa Wyższego, co pozwala na korzystanie z niej w bibliotekach uczelnianych lub poprzez *roaming access* z wykorzystaniem serwerów uczelnianych. Niezbędne jest jedynie aktywne konto w bibliotece uczelni.

Podobne funkcje pełnią jeszcze takie usługi, jak Scopus, Scirus czy Google Scholar.

W naukach medycznych natomiast niezwykle obszerną bazą zarówno danych bibliograficznych, jak i tekstów jest MEDLINE, stworzona i prowadzona przez Narodową Bibliotekę Medycyny Stanów Zjednoczonych. Baza dostępna jest poprzez trzech różnych dostawców: EBSCO, PubMed i ProQuest. Z kolei w naukach ścisłych wymienić należy na przykład arXiv – bazę danych bibliograficznych oraz artykułów opartą o filozofię *open access*³.

W ostatnich latach bardzo szybko rozwija się również nowy rodzaj baz danych bibliograficznych, tak zwane *crowdsourced database*, czyli baza danych tworzona i uzupełniana przez społeczność naukową, nie zaś bibliotekarzy czy automatyczne systemy katalogowania publikacji. Przykładem tego typu bazy może być Mendeley lub CiteUlike. Rozwiązania te stanowią nie tylko źródło danych bibliograficznych, ale dodatkowo pozwalają na organizowanie gromadzonej literatury i dzielenie się zebranymi bibliografiami. Pozwalają także śledzić, kto czyta publikacje, które są w naszym obszarze zainteresowań.

Po tym bardzo krótkim przedstawieniu wycinka dostępnych baz danych oraz usług pozwalających je przeszukiwać z wykorzystaniem Internetu, pozostaje pytanie, w jaki sposób ułożyć proces poszukiwania niezbędnej literatury i w jaki sposób organizować to, co uda nam się znaleźć.

³ Więcej baz danych działających w systemie *open access* można znaleźć na: http://en.wikipedia.org/wiki/List_of_open-access_journals.

Poszukiwanie i organizowanie literatury

To, w jaki sposób przebiega proces poszukiwania i gromadzenia literatury, w dużej mierze zależy od osobistych preferencji badacza i może być zależne od wcześniejszego doświadczenia, nawyków wyrobionych w trakcie prac badawczych, środowiska danej uczelni i współpracowników czy też po prostu wynikać z możliwości wykorzystania narzędzi, jakie zakupiła konkretna uczelnia. Uwzględniając wszystkie powyższe zastrzeżenia, można próbować wyodrębnić dwa schematy postępowania stosowane nierzadko równocześnie. Pierwszy polega na gromadzeniu danych poprzez konsultację z najbliższymi koleżankami i kolegami współdziałającymi zainteresowania badawcze, w dalszej części przejrzanie podstawowej literatury przedmiotu dostępnej „pod ręką”. W kolejnym kroku sprawdzenie katalogu przedmiotowego uczelnianej biblioteki, dalej przeszukanie zasobów indeksowanych przez Google Scholar lub wyżej wymienione usługi i bazy danych oraz odwiedzanie kolejnych stron wyszukiwarek. Drugi rozpoczyna się od przeszukania najbliższych zasobów, czyli plików znajdujących się w naszym komputerze, w naszych e-bookach, publikacjach w postaci PDF, lub stron WWW, także sprawdzonych do tego formatu, jak również plików i informacji uzyskanych od społeczności naukowej działającej w ramach *crowdsourced database*. Warto zadbać, aby te pozycje jako pierwsze zostały przez nas przejrzane.

Pozostaje zatem pytanie, w jaki sposób agregować zdobyte informacje, tak aby uniknąć duplikatów i powtórzeń oraz skrócić czas tworzenia prywat-

nej bazy danych bibliograficznych pod dany projekt badawczy, a dodatkowo przyspieszyć i usystematyzować pracę. Do tego celu możemy użyć oprogramowania wspierającego zarządzanie i organizację prywatnych baz bibliograficznych. Tego typu oprogramowanie funkcjonuje już dość długo – jednym z pierwszych programów, który nie tylko pozwalał na gromadzenie danych, ale także na ich późniejsze dostosowywanie do różnych stylów bibliograficznych, był BiBTeX, opracowany w latach osiemdziesiątych w Stanach Zjednoczonych. Zyskał on ogromną popularność i stał się protoplastą dla kolejnych produktów.

Oczywiście stosując to i jemu podobne oprogramowanie, badacz nadal może pozostać przy opisanym powyżej schemacie zdobywania danych, jednak tym razem zamiast przeklejać wyniki, jakie znajduje w poszczególnych bazach do „roboczego” pliku, może je po prostu automatycznie pobrać do RMS (Reference Management Software). Oprogramowanie automatycznie odczytuje metadane zawarte na stronie lub w PDF-ie i przyporządkowuje je do danego tytułu oraz uzupełnia prywatną bazę danych bibliograficznych o kolejne rekordy, wstawiając dane we właściwe pola i zgodnie z wybranym typem publikacji. Każdemu rekordowi zostaje nadany klucz, który pozwoli w późniejszym etapie tworzenia tekstu na łatwe odwołanie się do konkretnej publikacji z prywatnej bazy danych. Część programów RMS pozwala uniknąć odwiedzania kolejnych stron internetowych pozwalających na przeszukiwanie różnych baz danych, umożliwiając wyszukiwanie danych bibliograficznych i publikacji z głównego okna aplikacji po wcześniejszym skonfigurowa-

niu dostępności do, na przykład, płatnych serwisów. Dodatkowo, w przypadku wymienionego powyżej drugiego sposobu gromadzenia danych bibliograficznych oprogramowanie to pozwala nam na import wszystkich PDF-ów do programu, po czym może dokonać automatycznej segregacji, pobrania brakujących danych oraz zmian nazw, co pozwala uporządkować wszystkie pliki, na przykład po autorze czy dacie publikacji oraz oczywiście nadaje im indywidualne klucze niezbędne do wywoływania pozycji w tworzonym tekście. Niektóre z programów RMS pozwalają również na importowanie wszystkich komentarzy i notatek, jakie wykonane zostały na PDF-ach w czytnikach PDF-ów obsługujących taką funkcję. Pisząc o wspomnianej funkcji, należy zwrócić uwagę, że nie jest to już tylko zarządzanie danymi bibliograficznymi, ale także cytatai oraz – co równie ważne – wiedzą wytwarzaną przez badacza w trakcie gromadzenia, przeglądania i czytania publikacji.

Mówiąc o możliwościach gromadzenia danych bibliograficznych do projektu badawczego, należy również wspomnieć o pracy w zespole. Bez wykorzystania RMS bardzo często praca zespołowa ograniczała się do zgromadzenia przez poszczególnych członków zespołu bibliografii do przypisanych im rozdziałów. Praca nad tekstem także jest dzielona w ten sposób. Wykorzystanie technologii chmury obsługiwanej przez nowe programy RMS, ale także procesory tekstu i programy CAQDA, pozwala na zupełną zmianę sposobu pracy nad grupowymi projektami badawczymi.

Jednak mimo upływu czasu i rozwoju kolejnych wersji oprogramowań do organizacji i pracy z bi-

bliografią, wielu studentów i pracowników naukowych w Polsce i za granicą nadal nie korzysta z tego typu oprogramowania. Brak ten jest szczególnie widoczny w naukach społecznych. Dlatego w niniejszym artykule postaram się przekazać praktyczne informacje z zakresu organizowania projektu badawczego z wykorzystaniem nowych narzędzi RMS oraz przedstawić kilka wybranych menedżerów bibliografii wraz z ich podstawową funkcjonalnością. Należy jednak podkreślić, iż nie ma narzędzia, które w takim samym stopniu mogłoby być wykorzystywane do wielu różnych zadań, dlatego w trakcie omawiania autorskich propozycji, zostanie zwrócona uwaga na kompatybilność poszczególnych programów oraz ich funkcje dystynktywne.

Wśród wielu programów RMS do najbardziej popularnych należałoby zaliczyć: **Bookends**, **Citavi**, **Wkindx**, **Reference Manager**, **Referencer**, **Sente**, **Aigaion**, **KBibTeX**, **Mendeley**, **BibDesk**, **CiteULike**, **WizFolio**, **Papers**, **Biblioscope**, **Scholar's Aid**, **BibSonomy**, **EndNote**, **RefWorks**, **Bibus**, **Zotero**, **JabRef**, **RefDB**, **Qiqqa**, **Rebase**, **Pyblogger**, **Bebop**, **Jumper 2.0**, **Connotea**⁴.

Autor zdecydował się na bliższe przedstawienie jednego spośród wymienionych programów, a mianowicie CITAVI. Jest to bardzo dobre narzędzie do zarządzania danymi bibliograficznymi, PDF-ami, cytowaniami, komentarzami, umożliwia

⁴ Programy, których nazwy zostały zapisane pogrubioną czcionką, posiadają możliwość bezpośredniej współpracy z procesorami tekstu, takim jak np. MS Word. Z punktu widzenia usprawniania procesu tworzenia tekstów, raportów z badań czy opracowań teoretycznych, funkcjonalność ta ma ogromne znaczenie.

także współpracę zespołową czy też dzielenie się wynikami z szerszą społecznością naukową.

CITAVI jest programem płatnym, jednak jego licencja próbna pozwala na pracę przy pełnej funkcjonalności ze stoma rekordami. Zatem bez problemu można zrealizować mniejszy projekt badawczy z wykorzystaniem tego narzędzia. Pozostałe narzędzia, takie jak na przykład Mendeley oraz Zotero, są programami bezpłatnymi, które jednak całkowitą funkcjonalność uzyskują dopiero po zsynchronizowaniu z kontem zamieszczonym na serwerze, na którym użytkownik otrzymuje jedynie startową ilość miejsca, chyba że wniesie dodatkową opłatę. Zatem bezpłatna wersja wymienionych programów pozwala na realizację małych projektów i bardzo dobre poznanie programów, jednak w miarę rozwoju badacza-autora, zapotrzebowanie na przestrzeń serwerową wzrasta, a jej zwiększenie wymaga poniesienia dodatkowych opłat.

Citavi

Citavi jest projektem prowadzonym od 1995 roku, początkowo pod nazwą LiteRap, przez Swiss Academic Software na platformie Microsoft Windows. Do 2010 roku oprogramowanie to było rozwijane jedynie w języku niemieckim, jednak obecnie można korzystać z wersji wdrożonej w 2013 roku umożliwiającej obsługę programu w kilku językach, w tym w języku polskim.

Obecnie oprogramowanie oferowane jest w czterech wersjach:

- Citavi Free,
- Citavi Pro,

- Citavi Team,
- Citavi Reader.

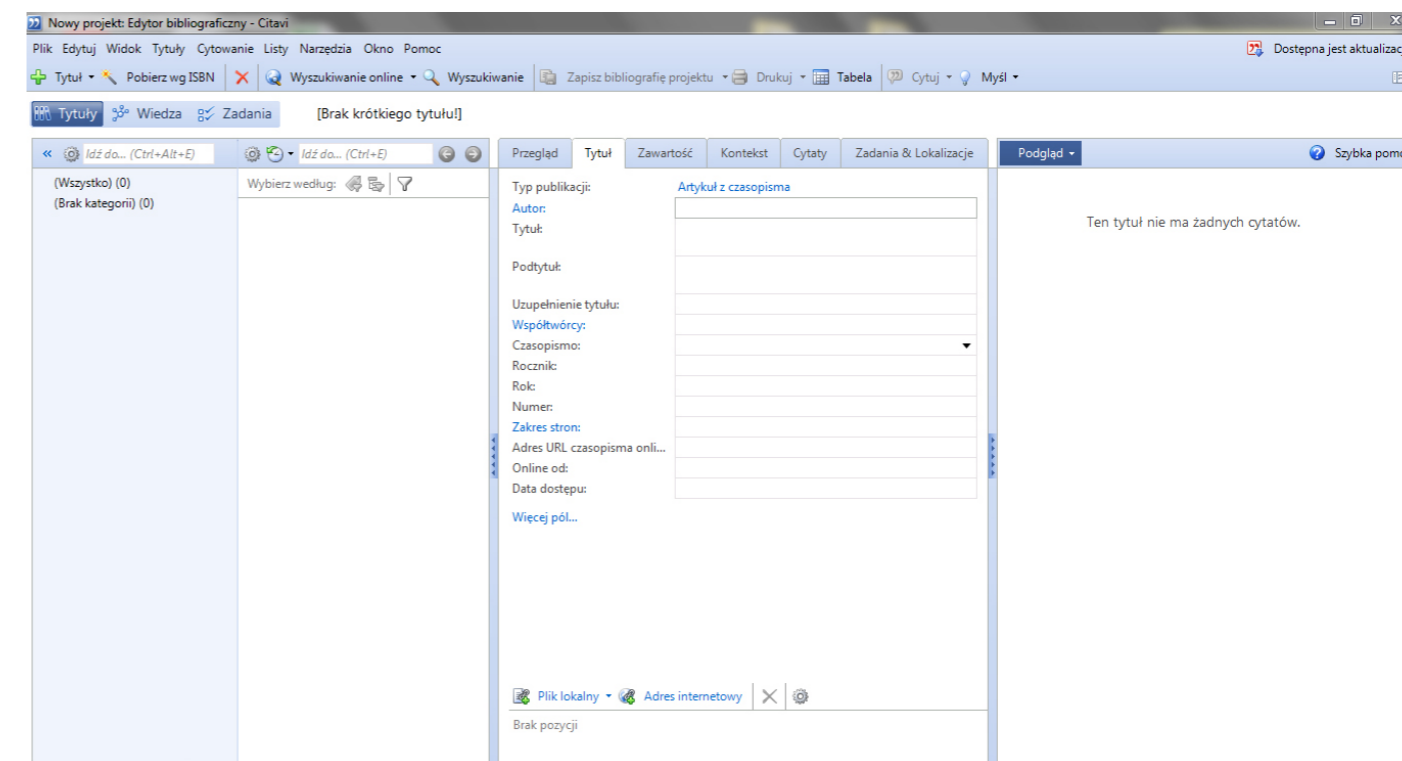
Citavi Free, jak już wspomniałem, umożliwia bezpłatne korzystanie z pełnej wersji programu w sposób nieograniczony czasowo. Jedyne ograniczenie polega na możliwości zarządzania maksymalnie stoma rekordami. Wersja Pro pozbawiona jest tych ograniczeń i może być używana komercyjnie⁵. Wersja Team programu pozwala na pracę zespołową, podczas gdy pozostałe wersje przeznaczone są dla projektów realizowanych przez pojedynczych badaczy⁶.

Wersję Citavi 4.2 Free możemy pobrać ze strony programu www.citavi.com. Po pobraniu pliku instalacyjnego, uruchamiamy instalację i w sposób standardowy przechodzimy przez kolejne kroki instalacyjne. Okno główne programu dzieli się na 4 podstawowe segmenty, które mogą być dowolnie przestawiane, włączane, wyłączane, w zależności od wykonywanej pracy. Filozofia interfejsu oraz jego strona graficzna oparte są na rozwiązaniach znanych z produktów firmy Microsoft, takich jak na przykład MS Word. Użytkownik pakietu Office nie powinien mieć więc problemu z intuicyjnością obsługi.

⁵ Studenci mogą zakupić program z 50% zniżką. Producent przewidział także specjalne oferty dla pracowników uczelni. Więcej na ten temat na stronie www.citavi.com.

⁶ Da się oczywiście korzystać z kilku osób z pliku programu w wersji Free lub Pro, należy jednak pamiętać, że korzystanie z nich przy jednoczesnym wykorzystaniu takich serwisów, jak Dropbox, służących do synchronizacji projektów prowadzonych na różnych komputerach, jest możliwe jedynie, jeśli są one zapisywane nierównoległe. Ostatnia wersja programu jest czytnikiem pozwalającym przeglądać projekty stworzone w pozostałych wersjach bez możliwości ich modyfikacji, chyba że nie przekraczają stu rekordów.

Rys. 4. Główny interfejs programu Citavi.



Źródło: opracowanie własne. Zrzut ekranu z programu Citavi 4.

W pierwszym ze wspomnianych segmentów pojawia się tworzona przez nas struktura zbieranego materiału w postaci drzewa kategorii. Dalej na prawo tworzony jest katalog publikacji dodawanych do programu, następny segment z zakładkami wyświetla wszystkie zgromadzone informacje dotyczące jednego tytułu. To w tym miejscu pojawiają się atrybuty z bazy danych bibliograficznych, które uzupełniane są automatycznie lub ręcznie – w zależności od sposobu ich pozyskania. Kolejne zakładki pozwalają gromadzić cytowania w różnych formach, a w ostatniej możemy zaplanować zadania do wykonania dotyczące danego tytułu oraz przypisać lokalizację całego tekstu, jeśli jest on dostępny w formie zdigitalizowanej. W ostatnim segmencie znajduje się podgląd danego tytułu: jeśli jest to strona WWW, wówczas możemy ją wyświetlić w tym

oknie, możliwe jest także przeglądanie plików PDF. Pracując na większym, panoramicznym monitorze, możemy jednocześnie wyświetlać wszystkie wymienione segmenty i w ten sposób łatwo tworzyć cytowania z czytanego tekstu, przyporządkowywać kategorie i słowa kluczowe oraz tworzyć komentarze czy notatki.

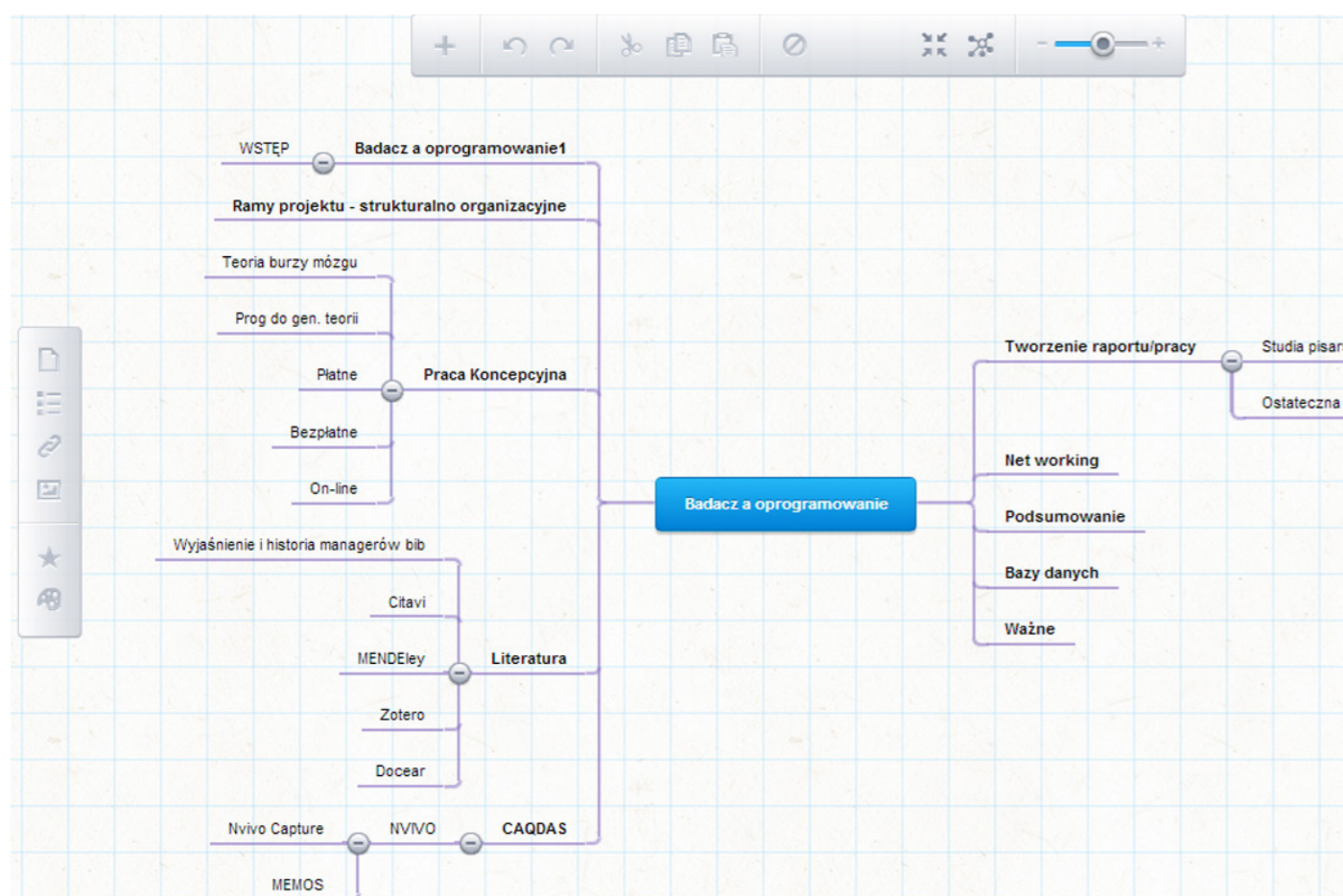
Praca z programem a praca nad projektem badawczym

Przejdźmy teraz do prześledzenia całego procesu pracy z programem, nie traktując jednak niniejszego artykułu jako szczegółowej instrukcji użytkownika, a jedynie wskazówkę do jego użytkowania i możliwości wykorzystania w pracy z innymi programami.

W celu wykorzystania pracy koncepcyjnej, jaką wykonaliśmy przy użyciu programu Mind42 lub Freemind (jeśli poddawaliśmy ją dalszej obróbce w tym programie) i nadania naszym poszukiwaniom literaturowym pewnego porządku oraz ram organizacyjnych, musimy naszą mapę myśli, repre-

zentującą w tym przypadku strukturę tworzonego artykułu, wyeksportować do formatu Freemind o rozszerzeniu .mm. Wykonanie tej czynności pozwoli nam zaimportować całą – wcześniej już opracowaną – strukturę do Citavi i umieścić ją w oknie reprezentującym kategorie.

Rys. 5. Mapa kategorii stworzonych w programie Mind42.

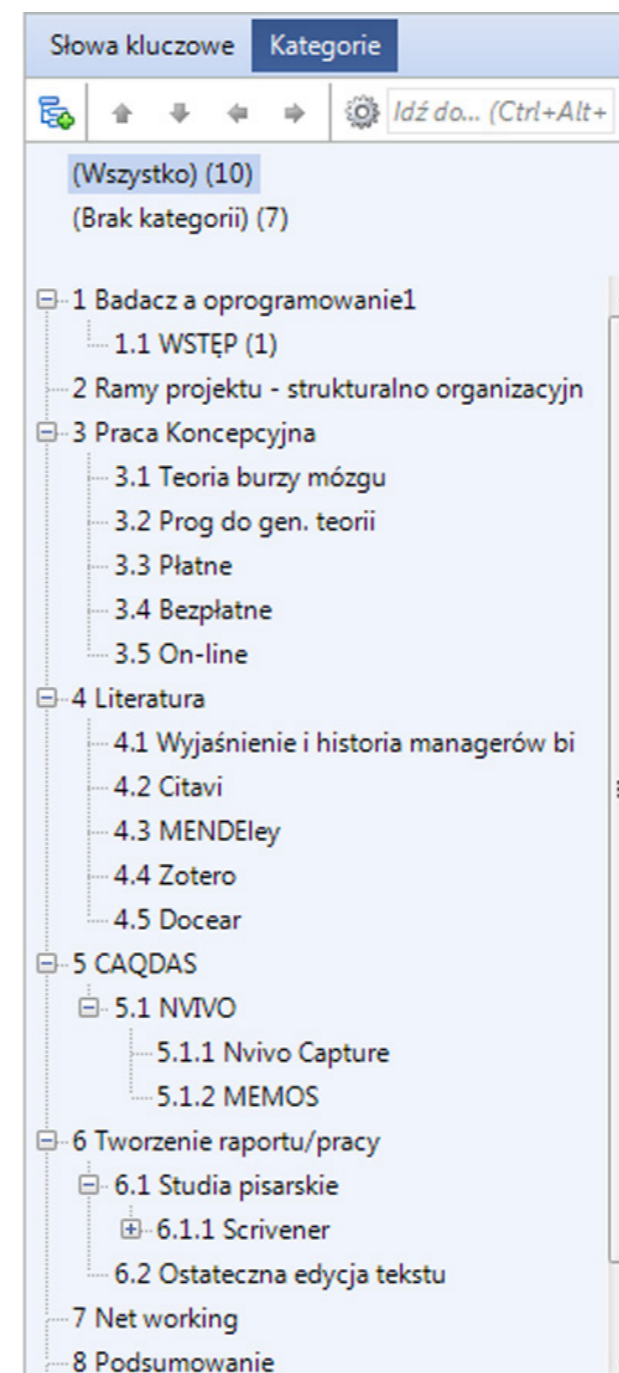


Źródło: opracowanie własne. Zrzut ekranu z programu Mind42.

Praca na opracowanych kategoriach pozwoli nadać rytm i skupić się na poszukiwaniu literatury z zarysowanego obszaru. Ponadto, obok słów kluczowych, które pozwalają panować nad rozrastającą się bazą danych, kategorie są jednym z bardzo ważnych elementów pracy. Służą one także do porządkowania cytowań, komentarzy i tworzonej

wokół nich bazy wiedzy. Warto także zauważyć, że w ramach Citavi możemy modyfikować nasze kategorie, dodawać nowe, usuwać je i porządkować. Następnie, przystępując do pracy nad konkretnym tekstem, możemy je eksportować do procesora tekstu czy studia pisarskiego i traktować jako konspekt naszej pracy.

Rys. 6. Drzewo kategorii zaimportowane z Mind42 do Citavi 4.



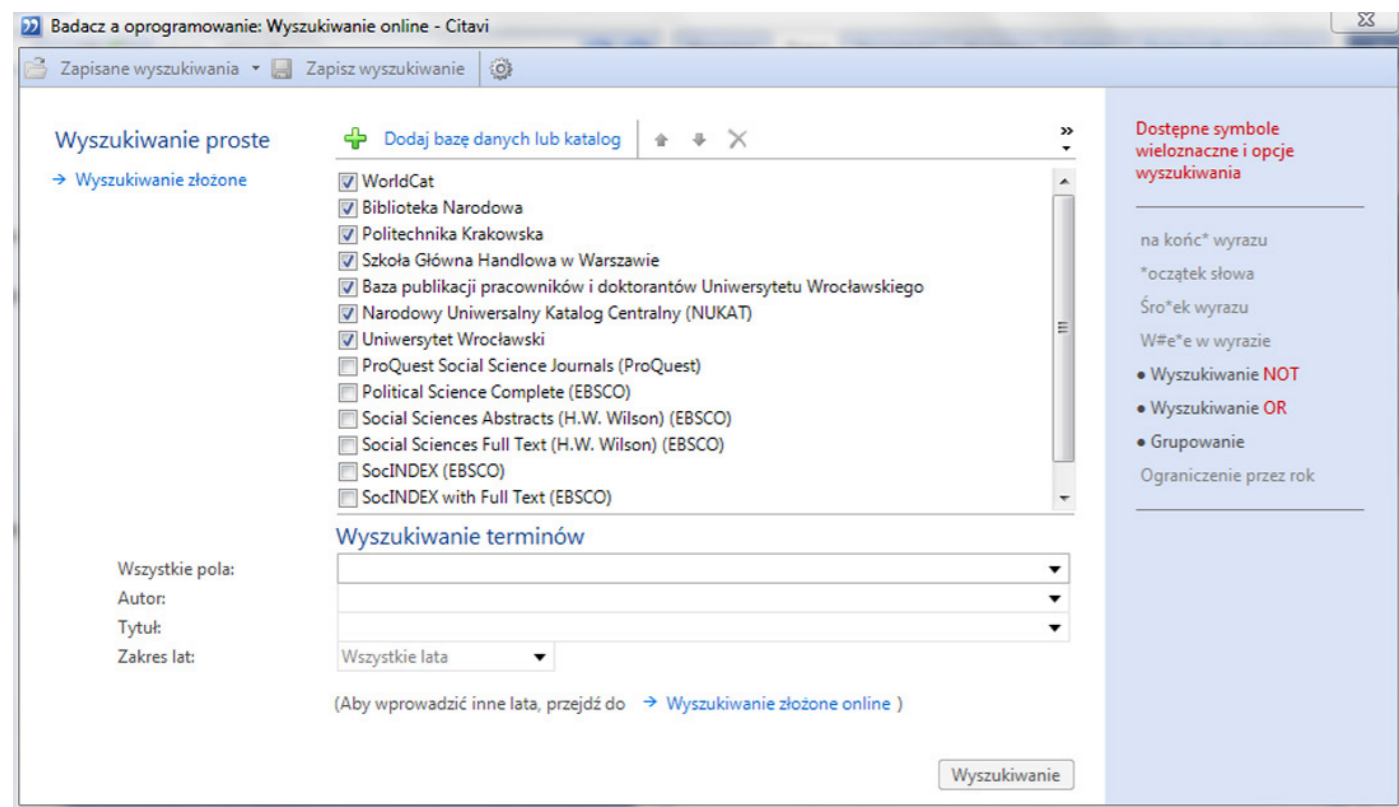
Źródło: opracowanie własne. Zrzut ekranu z programu Citavi 4.

Przechodząc dalej, po dokonaniu importu kategorii do Citavi, możemy skupić się na poszukiwaniach literatury, która będzie nam potrzebna przy realizacji naszego projektu badawczego.

W części dotyczącej poszukiwania danych bibliograficznych w bazach danych nadmieniałem, że programy RMS mogą wspierać badacza w tym zadaniu. Citavi pozwala na przeszukiwanie obecnie 4701 baz danych i katalogów bibliotecznych. Jeśli nasz komputer podłączony jest do sieci uczelnianej, a nasza uczelnia subskrybuje bazy danych, które chcemy przeszukać, bez problemu będziemy mogli wykonać zapytania do baz takich dostawców, jak EBSCO, OCLC, ProQuest, Web of Science oraz podobnych poprzez interfejs programu. Jeśli chcemy korzystać z możliwości wyszukiwania w subskrybowanych bazach z domu, program pozwala nam na konfigurację połączenia na przykład przez serwer naszej biblioteki. Zapamiętuje wówczas jego ustawienia, co pozwala na bezproblemowe wyszukiwanie poprzez tak zwany *roaming access*. Ponadto, w równie łatwy sposób przeszukamy WorldCat, jak i bazy biblioteki uczelnianej czy Biblioteki Narodowej. Obecnie producent dodaje systematycznie zgłaszane bazy biblioteczne do programu. Znajdziemy tam już bibliotekę na przykład SGH, Politechniki Krakowskiej, Uniwersytetu Wrocławskiego. Niestety nie ma tam katalogu Biblioteki Uniwersytetu Łódzkiego, jednak korzystając z katalogu NUKAT (Narodowy Uniwersalny Katalog Centralny), którego katalog Biblioteki Uniwersytetu Łódzkiego jest częścią, możemy przeszukiwać także zasoby biblioteki uniwersyteckiej. Warto pamiętać, że korzystając z wyszukiwania w bibliotece uczelnianej, Citavi może sprawdzić dostępność danej publikacji, pozwala także na wyszukiwanie publikacji poprzez numer ISBN, nazwę DOI lub PubMed Id.

Wyszukanie odpowiednich pozycji to pierwszy krok w tworzeniu prywatnej bazy danych na potrzeby realizowanego projektu. Kolejnym krokiem

Rys. 7. Okno wyszukiwania publikacji on-line w programie Citavi wraz z listą przeszukiwanych baz danych bibliograficznych.



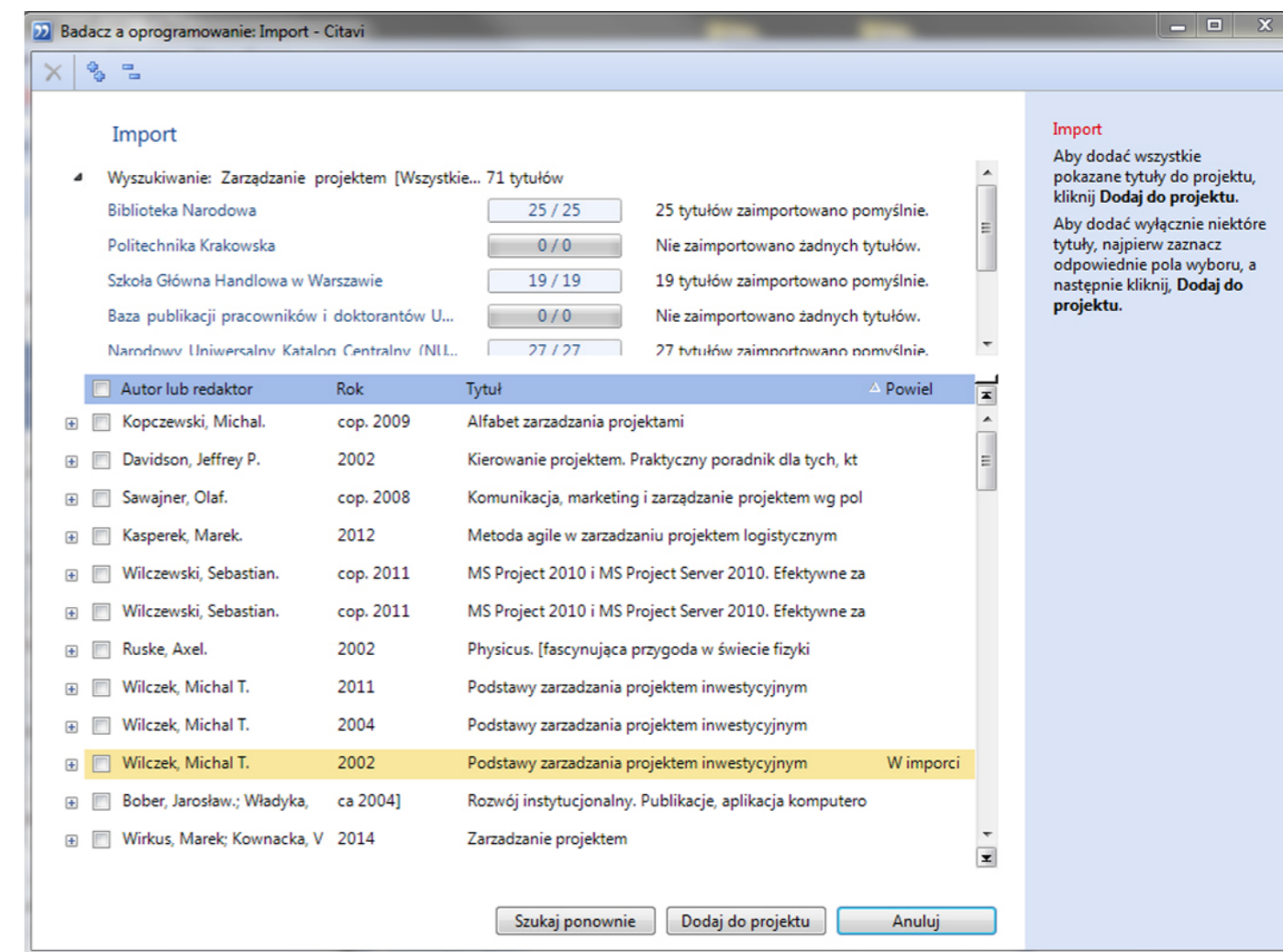
Źródło: opracowanie własne. Zrzut ekranu z programu Citavi 4.

jest zebranie wszystkich pozycji i dodanie ich do bazy, a następnie weryfikacja ich poprawności. Dodawanie rekordów wyszukanych poprzez interfejs programu jest bardzo wygodne i intuicyjne. Po wpisaniu zakresu wyszukiwania, program wyświetla listę publikacji spełniających wymagane kryteria.

Dodanie pozycji do projektu sprowadza się do zaznaczenia wybranych publikacji i wciśnięcia przycisku „Dodaj do projektu”. Wówczas program automatycznie doda pozycję do naszej bazy, uzupełniając wszystkie dostępne atrybuty, począwszy od podstawowych, takich jak autor, tytuł, miej-

sce publikacji, kończąc na spisie treści, abstrakcie i okładce, jeśli są dostępne. Innym sposobem budowania bazy danych jest wspomniane już dodawanie pozycji, na przykład poprzez numer ISBN. Jeśli chcemy przeszukać bazy niedostępne w opcji wyszukiwania on-line, wówczas możemy skorzystać z podprogramu, jaki instalowany jest razem z instalacją Citavi, czyli Citavi Picker. Ten niewielki program, będący dodatkiem do popularnych przeglądarek internetowych, takich jak na przykład Chrome czy FireFox, automatycznie odczytuje metadane zawarte na stronie internetowej oraz sugeruje pozycje, które można dołączyć do projektu. Jest to szczególnie przydatna funkcja

Rys. 8. Wyniki wyszukiwania publikacji w wybranych bazach danych.



Źródło: opracowanie własne. Zrzut ekranu z programu Citavi 4.

na przykład podczas przeglądania Google Scholar; gdy dana pozycja wydaje się godna uwagi, możemy ją w szybki sposób dodać do projektu i następnie poddać szczegółowej analizie. Dodatkowo jeśli przeglądamy bazy danych niedostępne przez port Z39.50, jak na przykład katalog Biblioteki Uniwersytetu Łódzkiego, wówczas wyszukując odpowiednią pozycję i otwierając widok szczegółowy, możemy również dodać ją bezpośrednio do projektu za pomocą Citavi Picker. Ikona programu re-

prezentowana jest jak na obrazie poniżej w postaci czerwonego znaku cudzysłowu.

Program ten umożliwia ponadto dodawanie do Citavi przeglądanych stron internetowych, PDF-ów, fotografii oraz zaznaczeń tekstu w postaci cytatów, co pozwala na rozszerzenie zarządzania zbieranymi informacjami o coraz bardziej popularne źródło, jakim jest Internet. Mając na uwadze wspomnianą funkcjonalność, należy wspomnieć,

Rys. 9. Okno katalogu bibliotecznego Biblioteki Uniwersytetu Łódzkiego wraz z wskaźnikiem Citavi Picker przy numerze ISBN.



Źródło: opracowanie własne. Zrzut ekranu z programu Citavi 4.

że zarządzaniem dokumentami internetowymi jeszcze lepiej radzi sobie Zotero, dlatego pracując głównie z tego typu dokumentami, warto wypróbować jego możliwości.

Wyszukiwanie pełnego tekstu

Funkcją, którą trudno przecenić w trakcie procesu gromadzenia tekstów i publikacji w pierwszej fazie projektu badawczego, jest możliwość wyszukiwania przez Citavi pełnego tekstu tytułu, którego

dane bibliograficzne zostały już dodane do bazy programu. Najbardziej dostępne są oczywiście artykuły z czasopism oraz gazet. Aby program wyszukał dostępne teksty dla całego projektu, należy z menu „Tytuły” wybrać opcję „Sprawdź dostępność i wyszukaj pełny tekst”, następnie wybrać „Znajdź pełny tekst” oraz zaznaczyć opcję „Pobierz pełny tekst”. Program przeszuka następujące bazy: WordCat, ArXiv, CrossRef.org, Digital Bibliography & Library Project, Directory of Open Access Journals, Google Patents, Pedocs, PubMed,

PubMed Central i dołączy do tytułów znalezione teksty, najczęściej w postaci plików PDF. Należy przy tym pamiętać, iż znacznie więcej trafień uzyskamy, jeśli pełnych tekstów będziemy poszukiwać po podłączeniu do sieci uczelnianej lub przez protokół VPN. Dzieje się tak, ponieważ część tekstów gromadzonych w takich bazach, jak na przykład EBSCO, dostępna jest jedynie po wykupieniu subskrypcji.

Integracja już posiadanych zasobów

Bazy danych i katalogi biblioteczne to nie jedyne miejsca, w których możemy odnaleźć interesujące nas publikacje. Z dużą dozą prawdopodobieństwa, jeśli kiedyś pracowaliśmy już nad obszarem, którego dotyczy nasz nowy projekt badawczy, to na naszym dysku znajdują się PDF-y lub inne pliki tekstowe zorganizowane w folderach systemowych, często posiadające nic nieznaczące nazwy. Aby nie stracić pracy, którą kiedyś wykonaliśmy, gromadząc pliki, można włączyć je do bazy danych tworzonej w Citavi. W tym zakresie program przewiduje dwie możliwości. Pierwsza, najprostsza, to przeciągnięcie plików z folderu otwartego w eksploratorze Windows do kolumny zawierającej tytuły w widoku głównym programu. Drugą metodą jest dodanie plików poprzez opcję „Import” wybraną z menu, co pozwala na wybranie folderu zawierającego pliki, które chcemy importować.

Wprowadzenie wszystkich plików, jakie udało nam się już wcześniej zgromadzić, do jednej bazy pozwoli uniknąć zbędnego duplikowania literatury. Mechanizmy wbudowane w Citavi weryfikują przed dodaniem nowych pozycji do bazy, czy dodawana

literatura już nie znajduje się w naszych zbiorach. Aby wykorzystać tę funkcję, bardzo ważne jest dokładne weryfikowanie rekordów uzupełnianych automatycznie i dopisywanie danych, które nie są zapisane w metadanych. Uchroni nas to przed powielaniem rekordów i jednocześnie zapewni poprawność i spójność naszej bibliografii przy zmianie stylów cytowań. Dla przykładu, jeśli będziemy w naszej bazie danych wpisywać tylko pierwsze litery imion autorów, może się to okazać niewystarczające w przypadku niektórych stylów cytowań. Takie braki w bazie nie pozwolą nam na swobodną zmianę stylów bibliograficznych, a tym samym zamiast usprawnić pracę, przysporzą dodatkowych problemów.

Kolejnym krokiem, który może być realizowany równoległe z gromadzeniem literatury, jest jej przegląd i analiza. Faza analizy często owocuje dużą liczbą notatek, komentarzy, nowych idei czy też cytatów, które mogą zostać wykorzystane przy opracowywaniu raportu lub publikacji. Wszystkie te elementy mogą być gromadzone w Citavi. Jeśli analizujemy literaturę dostępną nam jedynie w druku, cytaty, które chcemy wykorzystać, musimy przenieść ręcznie, przepisując treść lub z wykorzystaniem skanera i programów do rozpoznawania pisma zwanych OCR (Optical Character Recognition). W przypadku korzystania z publikacji występujących w formie elektronicznej, możemy cytować dowolne partie tekstu, wykorzystując opisany już powyżej Citavi Picker. Jeśli chcemy dodawać swoje uwagi i notatki do opracowywanych tytułów, możemy je również wykonywać bezpośrednio w Citavi. Pozwoli to w późniejszym czasie na łatwe zarządzanie cytowaniami i – poprzez

integrację z procesorem tekstu – znacznie przyspieszy pisanie publikacji. Ponadto, jeśli we właściwy i skrupulatny sposób będziemy przypisywać kategorie oraz słowa kluczowe do naszych cytatów, bez większego wysiłku będziemy mogli ich używać w innych publikacjach.

Etap 3 – zbieranie i analiza danych jakościowych

Analiza danych jakościowych – oprogramowanie CAQDAS

W trakcie prowadzenia projektu badawczego, po wykonaniu wstępnej pracy koncepcyjnej oraz pracy związanej z analizą dostępnej literatury (zgodnie

z naszym harmonogramem realizacji projektu, monitorowanym z wykorzystaniem ProjectLibre), przechodzimy do fazy zbierania danych i ich analizy. Wkraczamy wówczas w obszar analityczny, który może być wspomagany przez zastosowanie programu reprezentującego rodzinę CAQDAS (Computer Assisted Qualitative Data Analysis Software). Jeśli nasz projekt badawczy jest ograniczony w swych rozmiarach i stopniu zaawansowania, a my jako badacze rozpoczynamy dopiero przygodę z CAQDAS, warto dokładnie przyjrzeć się programom opartym o licencję GNU (bezpłatnym), które może nie są tak multifunkcjonalne jak programy na licencji komercyjnej, ale jednak w wielu przypadkach mogą okazać się wystarczające (Niedbalski 2013). Do rodziny programów CAQDAS zaliczamy między innymi:

Tabela 2. Zestawienie wybranych programów CAQDAS.

PROGRAM	DOSTĘPNOŚĆ	WPROWADZANE DANE
Coding Analysis Toolkit	bezpłatny (on-line)	tekst
QDA Miner Lite	bezpłatny – ograniczona funkcjonalność	tekst
RQDA	bezpłatny	tekst
Nvivo10	płatny	tekst oraz dane multimedialne
Atlas.ti7	płatny	tekst oraz dane multimedialne
QDA Miner	płatny	tekst
MAXQDA	płatny	tekst oraz dane multimedialne

Źródło: opracowanie własne.

W niniejszym artykule nie rozważam zaplecza historycznego programów CAQDAS, których początku możemy poszukiwać w oprogramowaniu powstającym w latach osiemdziesiątych, nie chcę także wskazywać na poszczególne zmiany zachodzące w kolejnych wersjach rozbudowywanego oprogramowania, o których więcej czytelnik może dowiedzieć się z takich źródeł, jak publikacje autorstwa Katarzyny Wilk (2001), Jacka Bielińskiego, Anny Rosińskiej-Kordasiewicz, Katarzyny Iwańskiej (2007), jak również Jakuba Niedbalskiego i Izabeli Ślęzak (2012).

Chciałbym natomiast dalszą część tego artykułu poświęcić krótkiej prezentacji możliwości jednego z przedstawicieli CAQDAS, jakim jest pakiet Nvivo w wersji 10, jako istotnego narzędzia w warsztacie badacza realizującego projekty badawcze zarówno na mniejszą skalę, jak i projekty międzynarodowe – wymagające pracy grupowej. Wybór tego pakietu jest arbitralny i wynika z doświadczeń i indywidualnych preferencji. Z pewnością część pojawiających się tu informacji może wydać się mniej dostępna dla osób, które nie pracowały z Nvivo, ale jednocześnie mam nadzieję, że pokazane narzędzia zachęcą do jego wypróbowania.

Nvivo 10 – zagadnienia wstępne

Na początku artykułu wskazałem – dla uproszczenia – że w projekcie badawczym możemy wyróżnić kilka jego faz, wymagających od badacza użycia różnych narzędzi wspomagających, takich jak choćby: programy do mapowania myśli, zarządzania bibliografią, organizowania wiedzy, zarządzania organizacyjną stroną projektu,

wspierające zbieranie danych i ich analizę, tworzenie raportów oraz publikacji. Teraz chciałbym zwrócić uwagę czytelnika na fakt, że etapowość projektu badawczego nie wyklucza korzystania z omawianych narzędzi w sposób równoległy i wzajemnie się uzupełniający. Dotyczy to szczególnie tak rozbudowanych programów, jak Nvivo. Pakiet ten nie ogranicza się obecnie jedynie do analizy danych, ale może być także środowiskiem ich agregowania, opracowywania, organizowania i wymiany. To, w jaki sposób Nvivo może wspierać badania prowadzone zgodnie z procedurą MTU, zostało już opisane w 2012 roku na łamach „Przeglądu Socjologii Jakościowej” w numerze 1 – w artykule *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej* autorstwa Jakuba Niedbalskiego i Izabeli Ślęzak. Spróbuję zatem, odnosząc się do tego artykułu, jak i do publikacji Schönfeldera (2011), przedstawić możliwe zastosowania Nvivo, wskazując raczej do czego badacz może wykorzystać to oprogramowanie, realizując projekt badawczy, niż podając szczegółowe instrukcje jego użytkowania.

Dobłą praktyką przy realizacji każdego projektu badawczego jest praca na wymienionych w artykule programach w kolejnych fazach opracowywania naszego projektu. Nvivo sprawdza się bardzo dobrze w zakresie gromadzenia danych, które będziemy z założenia poddawać dalszej obróbce, jaką jest na przykład kodowanie. Zatem tworzone mapy myśli w omawianym wcześniej programie Mindmap42 czy Freemind mogą być eksportowane do Nvivo, a wszelkie pomysły i koncepcje mogą być opisywane w formie notatek,

tak zwanych *memos*. Program umożliwia ponadto importowanie plików w postaci PDF wraz z poczynionymi adnotacjami i komentarzami, co znacznie ułatwia pracę, szczególnie jeśli naszym celem jest analizowanie zawartości importowanych pozycji. Nvivo sprawdza się również na etapie gromadzenia danych w postaci nagranych wywiadów, zdjęć, filmów, stron internetowych, wpisów na Facebooku, Twitterze, filmów i komentarzy na Youtube czy kwestionariuszy udostępnianych z pomocą serwisu SurveyMonkey. Program ten jest także nieocenionym wsparciem w procesie gromadzenia danych – dzięki wykorzystaniu podprogramu, jakim jest Ncapture – oraz przygotowania, opisu i organizacji danych przy wykorzystaniu linków, adnotacji, węzłów, klasyfikacji i kolekcji.

Analiza danych

Pracując nad projektem badawczym wykorzystującym dane jakościowe, wcześniej lub później musimy rozpocząć ich kodowanie. Korzystając z MTU, powiem o kodowaniu rzeczowym, które jest konceptualizacją treści empirycznej obszaru badanego i kodowaniu teoretycznym, które konceptualizuje relacje, w jakie kody rzeczowe mogą wchodzić i tworzyć hipotezy (Glaser, Strauss 1999; Konecki 2000; 2012; Charmaz 2009). Wspieranie kodowania, czyli nadawania partiom materiału kodów/etykiety wynikających z procesu analizy i interpretacji danych przez badacza, jest jednym z głównych celów oprogramowania CAQDAS, w tym także Nvivo. Jego twórcy dokładają wszelkich starań, aby ten proces był jak najbardziej efektywny i wygodny dla analityka. W pakiecie

Nvivo 10 zrezygnowano ze sztywnego tworzenia kodów/węzłów (*nodes*) zbieranych w sekcji *nodes* w folderach *free nodes*, *tree nodes* na rzecz dowolnego tworzenia folderów z nadawanymi przez analityka kategoriami kodów. Dzięki temu oprogramowanie to stało się jeszcze bardziej otwartym na prowadzenie projektów badawczych według różnych metodologii. To od badacza zależy więc, w jaki sposób będzie konstruował kody i wskazywał relacje pomiędzy nimi. Warto dodać, że korzystanie z programu wspierającego kodowanie pozwala z jednej strony być bardzo blisko danych, z drugiej zaś – wyabstrahować się od poszczególnych danych na rzecz ogólnego spojrzenia na kodowany materiał i tworzącą się strukturę kodów. Ponadto należy zwrócić uwagę na fakt, iż ręczne kodowanie nie wyczerpuje możliwości programu, a wręcz przeciwnie. W założeniu jego twórców udział kodowania ręcznego powinien być ograniczony do minimum, na rzecz coraz lepszych algorytmów automatycznego kodowania. Oczywiście takie spojrzenie otwiera pole do dyskusji, w której jest wiele stron oraz różnych argumentów, w dużej mierze zależnych od indywidualnych potrzeb badaczy. Dla niektórych automatyczne kodowanie jest nie do pomyslenia, inni są zainteresowani tak zwanymi *big data* i *mixed methods*, czyli pracą na potężnych zbiorach danych, liczących na przykład po kilka tysięcy wywiadów. Nvivo pozwala bowiem na ilościową analizę danych jakościowych dzięki takim narzędziom, jak choćby: zapytania (*queries*) o częstość występowania słów oraz ich rozkład w tekście, przeszukiwanie danych z wykorzystaniem synonimów słów, jak również specjalnie konstruowanych słowników czy auto-kodowanie.

Pozostając jeszcze przy pakiecie Nvivo, chciałbym zwrócić uwagę na dwie nowości, które w proponowanych dotychczas pozycjach z zakresu CAQDAS (oprócz Bazeley 2002) nie były poruszane. Dotyczą one gromadzenia materiałów do analizy z wykorzystaniem NCapture oraz eksperymentalnego modułu automatycznego kodowania z wykorzystaniem wzoru kodowego.

NCapture

NCapture jest osobnym programem rozwijanym przez firmę QSR, który jest instalowany razem z Nvivo 10 i stanowi rozszerzenie do przeglądarek internetowych, takich jak Internet Explorer i Chrome. Jego główną funkcjonalnością jest maksymalne uproszczenie i ułatwienie zbierania materiału badawczego dostępnego on-line, co jest szczególnie istotne w obecnych czasach, w których środowisko Internetu pozwala na prowadzenie badań w całości rozgrywających się w sieci (Miller 2012). Ponadto, a może przede wszystkim, jako badacze zbieramy także coraz więcej informacji uzupełniających nasze badania terenowe właśnie w Internecie, analizując strony internetowe dotyczące interesującego nas obszaru, przeszukując i analizując konwersacje prowadzone w mediach społecznościowych, takich jak chociażby Facebook czy analizując zawartość blogów lub filmów umieszczanych na Youtube.

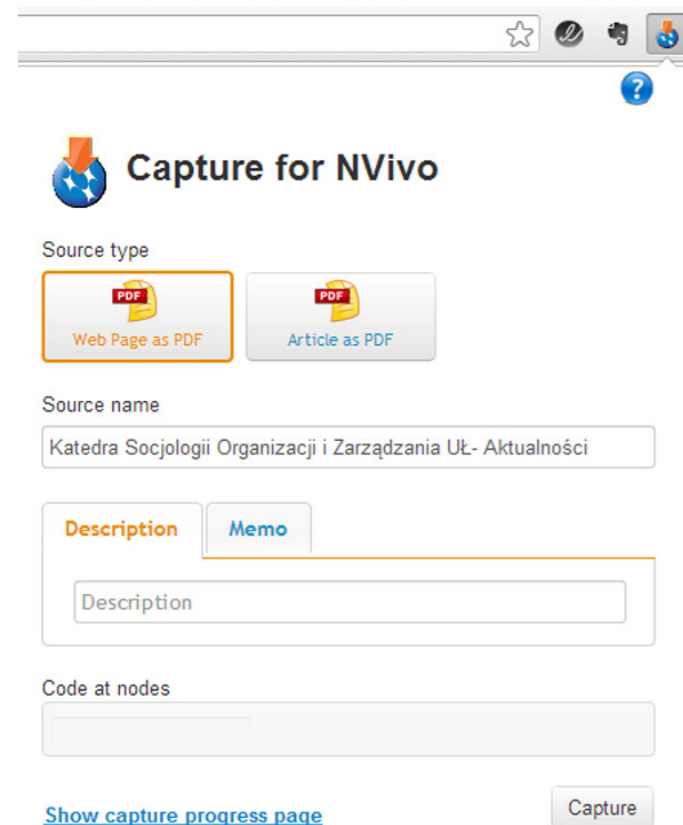
Przyjrzyjmy się zatem bliżej nowej propozycji firmy QSR. Po zainstalowaniu programu, w przeglądarce Chrome pojawia się w prawym górnym rogu przycisk, który odpowiada za uruchomienie NCapture. Jeśli uznamy, że przeglądana właśnie

strona WWW wymaga naszej uwagi i chcemy ją dodać do projektu badawczego prowadzonego w Nvivo, należy nacisnąć przycisk uruchamiający. Pojawi się wówczas interfejs programu, gdzie możemy wybrać typ źródła (*source type*), pod jakim zostanie ona zapisana. Do wyboru mamy dwie opcje. Pierwsza: *Web page as PDF* zapisuje całą stronę tak jak jest widoczna na ekranie. Jest to opcja, którą możemy wykorzystać, jeśli naszym celem badawczym jest na przykład analiza rozkładu treści wizualnych na stronach danego typu. Jeśli interesuje nas tylko treść artykułu, postu (np. na blogu), wybieramy opcję drugą – *Article as PDF*, zaprojektowaną z myślą o zbieraniu danych w postaci artykułów lub dłuższych postów wraz z komentarzami umieszczanymi na blogach. W kolejnym wierszu podajemy nazwę zapisywanego źródła. Dobrą praktyką jest przyjęcie i trzymanie się metodologii nazewnictwa źródeł od samego początku realizowania projektu badawczego i gromadzenia danych. Możemy w tym celu skorzystać ze skróconej nazwy naszego projektu badawczego, dalej posłużyć się datą i na przykład tytułem artykułu. W kolejnym kroku mamy możliwość umieszczenia opisu źródła (*description*), który pozwala uszczegółowić, jakiego rodzaju informacje zawiera zapisywana strona, w Nvivo będzie on widoczny jako właściwość źródła. Możemy także w chwili pobierania strony czy artykułu stworzyć notę (*memo*) wyjaśniającą, dlaczego zdecydowaliśmy się to źródło włączyć do projektu, jakie przesłanki nami kierowały i w jaki sposób dodawane dane odnoszą się zarówno do całości projektu badawczego, jak i do innych danych, które już zostały tam umieszczone. *Memo* będzie w Nvivo widoczna jako osobna notatka podlinkowana do

importowanego źródła. Dodatkowo, już w momencie dodawania źródła, możemy przypisać mu zupełnie nowy lub wcześniej używany w projekcie kod/węzeł (*Code at nodes*), jak również zakodować cały dokument w hierarchicznie stworzonych kodach, podając ścieżkę, na przykład: CAQDAS/nvivo/funkcje.

Po zaimportowaniu materiału do programu, kod zostanie przypisany całemu dokumentowi. Po uzupełnieniu wszystkich pól, co oczywiście nie jest wymagane, ale zdecydowanie pomaga organizować zbierany materiał, zatwierdzamy naszą decyzję przyciskiem *Capture*.

Rys. 10. Interfejs programu NCapture.



Źródło: opracowanie własne. Zrzut ekranu z programu NCapture.

Program zapisuje wskazany materiał i informuje nas o tym, czy proces przebiegł pomyślnie. Jeśli tak, spokojnie możemy przejść do kolejnego źródła. Nie ma potrzeby uruchamiania od razu Nvivo i importowania pobranego materiału. NCapture przechowuje bowiem zapisane dane w swoich katalogach i następnie pozwala przenieść je jednorazowo do Nvivo.

Jeśli duża część materiałów źródłowych opracowywanych przez nas jest dostępna na stronach internetowych w formacie PDF, to do ich importowania także możemy użyć NCapture. Choć PDF-y nie są tak kłopotliwe jak na przykład posty na blogach, gdyż możemy je w łatwy sposób zapisać na dysku twardym naszego komputera, a następnie dodać do Nvivo, to jednak i w tym przypadku możemy zyskać na czasie, korzystając z NC. Po pierwsze, przeszukując strony WWW, nie musimy za każdym razem zapisywać plików i wskazywać miejsca zapisu. Po drugie, jeśli import danych do Nvivo jest odroczone w czasie, trudno będzie nam sobie przypomnieć, dlaczego właśnie ten plik chcieliśmy dodać do projektu oraz jak dany plik wiąże się ze zbieranymi danymi. Wykorzystując NC, już w chwili dodawania dokumentu zapisujemy nasze uwagi w opisie, tworzymy *memo* z naszymi przemyśleniami powstałymi po szybkim przeskanowaniu dokumentu oraz kodujemy importowany materiał, umieszczając go tym samym w strukturze naszego projektu. Kolejnym źródłem danych są media społecznościowe. Ich szybki rozwój i wzrost użytkowników wymaga nowych narzędzi do prowadzenia badań społecznych w środowisku Web 2.0 lub z wykorzystaniem informacji zaczerpniętych z Internetu. Omawiany program pozwala pobierać dane z Facebooka, Twittera oraz LinkedIn. Procedura pozyskiwania postów czy komentarzy

z Facebooka jest bardzo zbliżona do tej opisanej przy stronach WWW. Jednak tym razem dodatkowo musimy być zalogowani na swoim koncie FB, a program będzie gotowy do pracy dopiero po udostępnieniu naszych danych publicznych z portalu. Domyślnie importowane dane w postaci postów czy komentarzy (jeśli importowane są z tablicy danego użytkownika, strony czy grupy) zapisywane są w postaci tabelarycznego zestawienia, które w późniejszej analizie niezwykle usprawnia ich przeglądanie, katalogowanie i analizowanie, jak również filtrowanie, kodowanie i przeszukiwanie. Program pozwala także grupować posty, na przykład według użytkowników, zakresu dat, lokalizacji i tym podobne, a uzyskany materiał można poddać dowolnym analizom, tak jak każdy materiał tekstowy. Jedynym ograniczeniem jest brak możliwości zapisu postów w postaci tabelarycznej, jeśli próbujemy importować dane z widoku strony głównej, gdzie pojawiają się posty umieszczone przez znajomych na ich tablicach. Wówczas mamy jedynie dostęp do zapisania widoku strony jako PDF. Dopiero po wejściu na dany profil lub ścianę grupy, mamy możliwość zapisu danych w wygodnej, tabelarycznej postaci. Ograniczenia narzędzia są w du-

żej mierze podyktowane ograniczeniami dostępu z naszego profilu do interesujących nas danych. Nie zawsze możemy zostać członkami zamkniętej grupy na FB, a nawet jeśli uda nam się uzyskać dostęp, musimy mieć świadomość, że niektóre grupy blokują programowi dostęp do danych poprzez ustawienia prywatności. Wówczas najlepiej skontaktować się z administratorem grupy.

Dane na facebookowych ścianach przyrastają nieustająco. Aby jednocześnie je zbierać i prowadzić analizę, możemy wybraną grupę na Facebooku wizualizować, wykorzystując dane do analizy klastrowej czy też analizować na mapie lokalizacje osób najczęściej komentujących. W chwili importowania danych NCapture każdorazowo będzie pobierał wszystkie dostępne posty w wybranej grupie. Jednak w trakcie importu do Nvivo, mamy możliwość scalenia wcześniejszych danych ze świeżo importowanymi, co pozwala na jednoczesne prowadzenie obserwacji i rejestrowanie aktywności w danej grupie oraz analizę zgromadzonego już materiału. Przykładowe dane po imporcie do Nvivo i ich wyczyszczeniu mogą wyglądać następująco.

Tabela 3. Przykładowe zestawienie tabelaryczne danych zaimportowanych z Facebooka.

POSTED BY (USERNAME)	POST	GENDER	COMMENTS USERNAME	COMMENT TEXT
Kowalski	Metody jakościowe mają nowe wsparcie w postaci kolejnej wersji programu Nvivo 10	male		
			Iksińska	Nie zapominajmy o MAXQDA :)
			Igrekowski	Ja dorzucam Atlas.ti7
			Głowacki	A ja zostaje przy kartkach długopisie i nożyczkach :)
Nowak	Przydałyby się jakieś fajne warsztaty z CAQDAS :)	female		

Źródło: opracowanie własne.

Pobieranie danych z Twittera, czyli tak zwanych tweetów, rozpoczyna się również od autoryzacji aplikacji. Podobnie jak w przypadku Facebooka, tak i z Twittera możemy zapisywać tweety z naszej linii czasu tylko w postaci PDF-ów. Jako dane tabelaryczne (*datasets*) możemy natomiast pobierać tweety osób lub organizacji, które śledzimy. W zestawieniach tabelarycznych mamy także możliwość zapisywać tweety, które pokazują się jako wyniki wyszukiwania, na przykład możemy poprzez Twittera wyszukać wszystkie tweety oznaczone hasztagiem CAQDAS (#CAQDAS) i zapisać je w zestawieniu, które po zaimportowaniu do Nvivo pozwoli nam na dalszą analizę i kodowanie. Podobnie jak przy importowaniu danych z Facebooka, tak i w tym przypadku zestawienia tabelaryczne podlegają wszystkim możliwym w Nvivo przekształceniom.

Pobieranie danych z serwisu LinkedIn podlega takim zasadom, jakie zostały opisane w przypadku Facebooka i Twittera. W wypadku tej społeczności wyszukujemy interesującą nas grupę i możemy zapisywać trwające w niej dyskusje lub wybrać tę dyskusję, która najbardziej nas interesuje. W związku z tym, że LinkedIn wprowadziłienne limity pobierania danych, które kasuje się o północy czasu UTC, importując do Nvivo wszystkie dyskusje w dużej grupie, możemy wyczerpać limit pobrania danych, zanim NCapture zapisze wszystkie wątki i dyskusje. Należy więc zwrócić na to szczególną uwagę, zanim przystąpimy do pobierania danych.

W związku z wykorzystywaniem w badaniach jakościowych w coraz większym zakresie analizy nagrań wideo, firma QSR wprowadziła do NCapture funkcjonalność bardzo ułatwiającą pozyskiwanie

i organizację nagrań znajdujących się w serwisie Youtube. Pobranie filmów nie wymaga, tak jak we wcześniejszych przypadkach, autoryzacji ani też zalogowania do serwisu, pod warunkiem niepobierania wideo oznaczonych jako prywatne. Import filmów przebiega poprzez znany nam już interfejs. W tym przypadku mamy do wyboru trzy możliwe typy źródeł (*source types*): tylko wideo, wideo i komentarze oraz strona jako PDF. W pierwszym przypadku program pobierze link do oryginału. Po zaimportowaniu filmu do Nvivo, zostanie on umieszczony w folderze źródła, skąd będziemy mogli go odtwarzać, aby dokonywać transkrypcji czy kodowania. Jeśli wybierzemy opcję pobrania wraz z komentarzami, wówczas film zostanie w trakcie importu do Nvivo umieszczony tak jak w poprzedniej opcji, natomiast komentarze zostaną zapisane jako znane nam zestawienie tabelaryczne (*dataset*). Niestety NCapture zapisuje i importuje jedynie link do oryginału, co z jednej strony pozwala na utrzymywanie mniejszej bazy danych nieobciążonej nagraniami wideo dostępnymi na serwerach Youtube, z drugiej jednak – w chwili usunięcia pliku z serwisu tracimy do niego dostęp. Zatem zaproponowana przez QSR funkcja niestety nie zastąpi nam oprogramowania do pobierania oryginalnych plików i zapisywania ich na dysku, do czego możemy używać na przykład bezpłatnego Vdownloadera.

Zagadnienia do dyskusji

Kończąc temat prezentacji narzędzia, jakim jest oprogramowanie Nvivo, chciałbym jeszcze zwrócić uwagę czytelnika na kilka zagadnień, które w moim odczuciu nie są jeszcze wystarczająco eksplorowane w polskiej literaturze przedmiotu.

Zagadnieniem kluczowym dla rozwoju każdego obszaru badań i metodologii jest pytanie o sposób nauczania prowadzenia badań jakościowych z wykorzystaniem nowych technologii, w tym oprogramowania CAQDAS. Innymi słowy, jest to pytanie o to, czy przyszli badacze otrzymują wystarczające wsparcie i pomoc we właściwym rozumieniu funkcji i ograniczeń oprogramowania CAQDAS. Co za tym idzie, pojawia się pytanie o dwie kolejne kwestie. Po pierwsze, podejście dydaktyków na polskich uczelniach do oprogramowania komputerowego wspierającego prowadzenie badań jakościowych, po drugie – sposób, w jaki podręczniki metodologiczne wprowadzają adeptów badań społecznych w świat prowadzenia badań z użyciem oprogramowania CAQDAS. Jest to zagadnienie bardzo szerokie i poznawczo niezwykle ciekawe. Wystarczy bowiem dla przykładu zwrócić się w stronę jednego z bardziej popularnych podręczników metodologicznych, jakim jest *Prowadzenie badań jakościowych* Davida Silvermana (2009), aby zobaczyć, że główną zaletą wymienianą przez autora jest „szybkie opracowywanie dużych zbiorów danych, co uwalnia badacza od rozwiązywania licznych problemów analitycznych” (s. 234). Takie postrzeganie programów CAQDAS w pewnym stopniu nasuwa skojarzenia i chęć porównywania tego oprogramowania z oprogramowaniem do ilościowej analizy danych. Musimy jednak zwrócić uwagę na to, iż zastosowanie siatki poznawczej przeniesionej z bardziej popularnego oprogramowania ilościowego, takiego jak SPSS czy STATISTICA, prowadzi do zafałszowania użyteczności CAQDAS i nie pozwala w pełni dostrzec i zrozumieć filozofii ukrytej w tych programach (Niedbalski 2013). Niewłaściwe jest także postrzeganie

CAQDAS jako oprogramowania, które umożliwia jedynie ilościową analizę danych jakościowych (Davidson, Skinner 2010). Musimy bowiem pamiętać, że wprowadzanie danych jakościowych do programu już jest ich interpretacją (Bazeley 2012).

Kolejnym zagadnieniem wynikającym poniekąd z posługiwania się siatką pojęć zaczerpniętą z badań ilościowych jest zasadność pytania o to, czy analizy wykonane z użyciem oprogramowania CAQDAS są bardziej przejrzyste, rygorystyczne, powtarzalne i godne zaufania. Tego typu rozważania pojawiają się w literaturze przedmiotu (Smyth 2008), jednak w tym artykule chciałbym podkreślić, że prezentowane oprogramowanie wspierające warsztat badacza nie zastąpi jego umiejętności analitycznych i wiedzy metodologicznej oraz nie sprawi, iż jego analizy – tylko dzięki użyciu oprogramowania – staną się lepsze, bardziej odkrywcze czy też teoria opracowana na ich podstawie będzie lepiej pracować. Nie zmienia to jednak faktu, że wprowadzenie do analizy jakościowej algorytmów wykorzystywanych przez oprogramowanie z rodziny CAQDAS pozostawia otwartym pytanie o wpływ tych programów na proces analizy oraz samo doświadczanie tego procesu przez badacza (Davidson, Skinner 2010).

Warto także refleksyjnie spojrzeć na zagadnienie związane z wpływaniem oprogramowania na wybór metodologii oraz kierunkowanie analizy, jak również etykę badań jakościowych prowadzonych z użyciem CAQDAS. Nowe możliwości takich pakietów, jak Nvivo, Atlas.ti pozwalają na ominięcie etapu transkrypcji, co z jednej strony zmniejsza straty w danych (pozwalając na kodowanie bezpośrednio na wywiadzie), z drugiej – stawia pytanie

o sposób chronienia anonimowości i poufności badanych (praca z innymi badaczami nie odbywa się na transkrypcji, ale materiale audio/wideo) (Evers 2011). Ponadto ważką kwestią staje się bezpieczeństwo danych przy pracy w dużych zespołach z wykorzystaniem technologii chmury.

Pojawia się także pytanie bardziej prozaiczne i techniczne, a mianowicie – do czego tak właściwie mogą zostać użyte programy z rodziny CAQDAS, jaka jest ich funkcjonalność i jakie są ich cechy dystynktywne (Di Gregorio 2000; Bazeley 2002; Hutchison, Johnston, Breckon 2010; Niedbalski, Ślęzak 2012). Odpowiedź na to pytanie jest jednocześnie odpowiedzią, który program (jeśli w ogóle) wybrać do naszego warsztatu i w oparciu o jakie przesłanki powinniśmy dokonać tego wyboru. W niniejszym artykule pozostanę przy stwierdzeniu, iż zawsze ostateczna decyzja, czy i który program wybrać, pozostaje w domenie indywidualnego badacza. Decyzja, jaki program zostanie wybrany do prowadzenia dalszej analizy gromadzonego materiału i czy będzie to program z rodziny CAQDAS, powinna być podyktowana przede wszystkim rozsądkiem, choć w dużej mierze może także zależeć od czynników tak prozaicznych, jak choćby dostępność danego programu na uczelni. Jeśli projekt jest niewielki, a analizowane dane nie są bardzo obszerne i różnorodne, można zamiast rozbudowanego programu wybrać program do tworzenia map konceptualnych lub prosty program do kodowania, jakim jest OpenCode, pamiętając, że oprogramowanie to wspiera metodologię teorii ugruntowanej (Niedbalski 2013). Jeśli jednak materiał do analizy jest bardzo obszerny i wymaga integracji różnych rodzajów danych, warto zastanowić się, czy nie włączyć do swojego warsztatu jednego

z wymienionych przedstawicieli rodziny CAQDAS. Jest to tym bardziej zasadne, jeśli będziemy chcieli używać w projekcie metod mieszanych i dokonywać ilościowej analizy danych jakościowych. Jeśli jednak nasz problem badawczy oraz wybrana metodologia skupiają się na przykład wokół analizy dyskursu, to warto wówczas rozważyć takie pakiety oprogramowania, jak na przykład Leximancer, który wspomaga analizę treści i tworzenie map koncepcyjnych.

Chciałbym zaznaczyć raz jeszcze, iż Nvivo nie jest jedynym wyborem, szczególnie jeśli dopiero rozpoczynamy przygodę z CAQDAS. Warto bowiem rozważyć narzędzia, które dostępne są bezpłatnie i mimo że nie zawsze są tak rozbudowane jak płatne programy, to może się okazać, iż będą dla nas wystarczające. W tym miejscu polecam zapoznanie się z już wspomnianą publikacją *Odkrywanie CAQDAS* (Niedbalski 2013), w której znajdziemy opisy takich programów, jak OpenCode, Weft QDA, CmapTools, EZ-text, TranscriberAG. Zachęcam do próbowania i testowania omawianych programów, być może któryś z nich stanie się głównym narzędziem wspomagającym realizowane przez nas projekty badawcze.

Etap 4 – upowszechnianie wyników w formie publikacji

Tworzenie raportu/pracy

Po zebraniu materiału badawczego, poddaniu go analizie, zebraniu na nowo, porównaniu i ponownym przeanalizowaniu, dochodzimy do momentu, w którym musimy zebrać wszystkie nasze wyniki, przemyślenia, spostrzeżenia, hipotezy, załączki teo-

rii, a może nawet teorie i przedstawić je szerszej społeczności (zarówno naukowej, jak i każdemu, kto mógłby być zainteresowany przebiegiem naszych badań i ich wynikami). W naukach społecznych jedną z form prezentacji wyników jest raport badawczy, a w dalszej kolejności publikacja. Jeśli nasz projekt był realizowany jako praca magisterska czy doktorska, tym bardziej opracowanie jego wyników powinno przyjąć formę dobrze zorganizowanego tekstu prezentującego nasze założenia, realizację i wyniki.

Programy wspierające pracę z tekstem możemy podzielić na programy (tak zwane studia pisarskie), których przeznaczeniem jest wspieranie autora w tworzeniu bardziej rozbudowanych publikacji, takich jak książki, scenariusze, raporty oraz na programy, które są procesorami tekstu, zatem umożliwiają jego wprowadzanie i skład. Ze względu na podejście do składu dokumentu, możemy wyróżnić programy obsługujące formatowanie wizualne (WYSIWYG – What You See is What You Get) i programy obsługujące formatowanie logiczne (WYSIWYW – What You See Is What You Want).

Tabela 4. Zestawienie wybranych programów do pracy z tekstem.

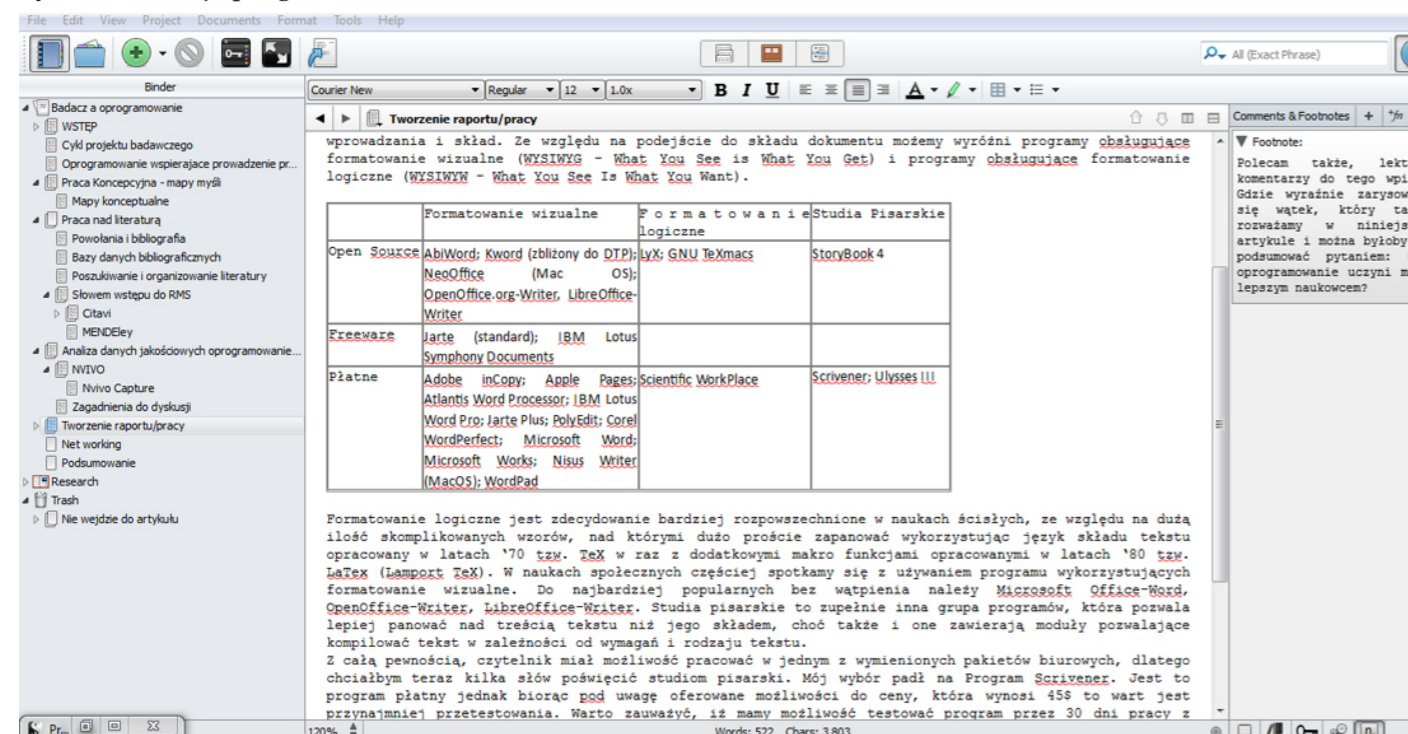
DOSTĘPNOŚĆ	FORMATOWANIE WIZUALNE	FORMATOWANIE LOGICZNE	STUDIA PISARSKIE
<i>open source</i>	AbiWord, Kword (zbliżony do DTP), NeoOffice (Mac OS), OpenOffice Writer, LibreOffice Writer	LyX, GNU TeXmacs	Story Book 4
<i>freeware</i>	Jarte (standard), IBM Lotus Symphony Documents		
płatne	Adobe InCopy, Apple Pages, Atlantis Word Processor, IBM Lotus Word Pro, Jarte Plus, PolyEdit, Corel Word Perfect, Microsoft Word, Microsoft Works, Nisus Writer (Mac OS), WordPad	Scientific WorkPlace	Scrivener, Ulysses III

Źródło: opracowanie własne.

Formatowanie logiczne jest zdecydowanie bardziej rozpowszechnione w naukach ścisłych, ze względu na dużą liczbę skomplikowanych wzorów, nad którymi dużo prościej zapanować, wykorzystując język składu tekstu opracowany w latach siedemdziesiątych, tak zwany TeX, wraz z dodatkowymi

makrofunkcjami opracowanymi w latach osiemdziesiątych, tak zwanymi LaTeX (Lamport TeX). W naukach społecznych częściej spotkamy się z używaniem programów wykorzystujących formatowanie wizualne. Do najbardziej popularnych bez wątpienia należy Microsoft Office Word, OpenOffice

Rys. 11. Interfejs programu Scrivener.



Źródło: opracowanie własne. Zrzut ekranu z programu Scrivener.

Writer, LibreOffice Writer. Z całą pewnością czytelnik miał możliwość pracować z jednym z wymienionych pakietów biurowych, dlatego chciałbym teraz kilka słów poświęcić studiom pisarskim, stanowiącym zupełnie inną grupę programów pozwalających lepiej panować nad treścią tekstu niż jego składem, choć zawierają również moduły pozwalające kompilować tekst w zależności od wymagań i rodzaju tekstu. Bliższego opisu tej grupy programów dokonam na przykładzie programu Scrivener.

Jest to program płatny, jednak biorąc pod uwagę stosunek oferowanych możliwości do ceny, która wynosi 45 dolarów, to wart jest przynajmniej przetestowania, które może trwać do 30 dni⁷. Nie chciałbym w tym

⁷ W przypadku tego oprogramowania jest mowa o trzydziestu dniach roboczych. Oznacza to, że korzystając z programu dwa dni w tygodniu, możemy go bez przeszkód testować przez kilkanaście tygodni.

miejsu prezentować szczegółowego opisu programu, który czytelnik może odnaleźć na blogu Emanuela Kulczyckiego w poście pod tytułem „Scrivener – najlepszy program do pisania artykułów i książek”⁸, zwrócę jednak uwagę na kilka funkcji, jakie w bardzo dużym stopniu mogą ułatwić tworzenie tekstu publikacji w oparciu o noty sporządzone w Nvivo.

Scrivener prezentuje zupełnie inne podejście do pracy z tekstem. Na próżno szukać tu białej kartki takiej jak w MS Word. Główny interfejs dzieli się na trzy segmenty, w których mamy stały podgląd struktury naszego dokumentu, aktualnie wprowadzany tekst oraz – zupełnie po lewej stronie notatki – komentarze i przypisy.

⁸ Polecam także lekturę komentarzy do tego wpisu, w których wyraźnie zarysowuje się wątek, jaki również rozważam w niniejszym artykule, i można byłoby go podsumować pytaniem: czy oprogramowanie uczyni mnie lepszym naukowcem?

Do kolumny Binder możemy zaimportować strukturę naszego tekstu, którą przygotowaliśmy w fazie koncepcyjnej pracy nad projektem, korzystając na przykład z Mind42 lub Freemind. Możemy ją także eksportować po zakończeniu pracy do formatu .doc lub .mm i użyć jako mapy. Każda z sekcji reprezentowana w strukturze może być traktowana jako osobny dokument. Elementy tam umieszczone możemy dowolnie przedstawiać, wykorzystując do tego widok tablicy korkowej. Z punktu widzenia integrowania not tworzonych w Nvivo, opisana filozofia traktowania sekcji jako oddzielnych plików jest nieoceniona. Możemy bowiem do programu importować wszystkie nasze noty wcześniej zapisane do formatu .doc. Pojawią się one wówczas jako odrębne sekcje w kolumnie Binder. Następnie możemy dokonywać dowolnych zmian i korekt, możemy je układać, przeklejać, organizować i formatować, po czym w prosty sposób złożyć do jednego dokumentu i zapisać w wybranym formacie, takim jak na przykład .doc, .odt czy .rtf. Mimo że osobom przyzwyczajonym do procesorów tekstu takich jak Word trudno jest oswoić się z inną filozofią organizacji tekstu, to z całą pewnością wysiłek włożony w naukę nowego programu wynagradza pełnoekranowy tryb pisania tekstu z funkcją środkowania, dzięki czemu nigdy nie piszemy na dole ekranu. Dodatkowo, Scrivener oferuje nam statystyki użytych wyrazów, pozwala ustalać limity znaków oraz dzienne limity znaków pozwalające na utrzymanie tempa pracy, które wcześniej sobie założyliśmy. Końcowe składanie tekstu bardziej przypomina logikę What You See Is What You Want, gdyż nie formatujemy tekstu tak jak w MS Word, a jedynie tworzymy maskę, która pozwala sformatować dokument dokładnie tak, jak zostanie to określone. Zajmuje to trochę czasu, jednak utworzonego w ten sposób schematu można następnie używać jako layoutu, na

przykład pod konkretne czasopismo. Jedną z ważnych cech tego oprogramowania jest także bezpieczeństwo naszych danych. Program automatycznie zapisuje naszą pracę w czasie rzeczywistym, a przy zamykaniu programu tworzy kopię bezpieczeństwa we wcześniej wskazanej lokalizacji.

Podsumowanie

Podobnie jak społeczeństwo, tak i obszary, metody i narzędzia badań społecznych ulegają zmianom. Oczywiście można prowadzić świetne projekty badawcze, posługując się ołówkiem i papierem, nie mniej jednak wykorzystanie nowych narzędzi wspierających pracę koncepcyjną i analityczną może przyczynić się do zwiększenia efektywności badań. Programy do zarządzania projektem badawczym pozwalają utrzymać rytm pracy, zwiększając kontrolę oraz pozwalają bardziej efektywnie realizować większe projekty. Menedżery bibliografii oraz cytowań pomagają lepiej panować nad wykorzystywaną literaturą oraz wspierają tworzenie i weryfikację wiedzy przez społeczność naukowców skupionych wokół takich mediów społecznych, jak Mendeley czy LinkedIn. Nowe generacje oprogramowania CAQDAS dają możliwości zbierania coraz bardziej różnorodnych danych jakościowych chociażby z takich źródeł, jak Facebook czy Twitter. Ponadto rozwijane możliwości kodowania na materiale źródłowym bez potrzeby transkrypcji, choć rodzą problemy etyczne, to realnie przyspieszają proces badawczy. Postępująca integracja pomiędzy poszczególnymi programami pozwala z kolei na bardziej płynne przechodzenie przez poszczególne etapy procesu badawczego oraz właściwy przepływ informacji zarówno w wewnętrznym komputerowym środowisku pracy każdego badacza, jak

i pomiędzy członkami zespołu badawczego czy też różnymi zespołami. Sądzę więc, że warto poświęcić trochę czasu na przetestowanie różnych możliwości organizacji swojego warsztatu pracy, tak aby móc maksymalnie zoptymalizować narzędzia, którymi się

posługujemy przy realizacji projektów badawczych. Pamiętajmy jednak, że zaproponowana przeze mnie kompilacja programów oraz aplikacji sieciowych to jedynie ułamek możliwości, jakimi w społeczeństwie 2.0 dysponuje badacz.

Bibliografia

A Guide to the Project Management Body of Knowledge (PMBOK guide) – Fifth Edition (2013) Newtown Square, PA: Project Management Institute [dostęp 13 maja 2013 r.]. Dostępny w Internecie: http://www.projectauditors.com/Papers/PMBOK_FifthEdition.pdf.

Aronson Elliot, red., (2001) *Człowiek istota społeczna*. Przełożyła Lidia Grzesiuk. Warszawa: Wydawnictwo Naukowe PWN.

Bazeley Patricia (2002) *The Evolution of a Project Involving an Integrated Analysis of Structured Qualitative and Quantitative Data: From N3 to NVivo*. „International Journal of Social Research Methodology”, vol. 5, no. 3, s. 229–243.

----- (2012) *Regulating Qualitative Coding Using QDAS? „Sociological Methodology”*, vol. 42, no.1, s. 77–78.

Charmaz Kathy (2009) *Teoria ugruntowana. Praktyczny przewodnik po analizie jakościowej*. Przełożyła Barbara Komorowska. Warszawa: Wydawnictwo Naukowe PWN.

Davidson Lorraine, Skinner Heather (2010) *I Spy with My Little Eye: A Comparison of Manual versus Computer-aided Analysis of Data Gathered by Projective Techniques*. „Qualitative Market Research: An International Journal”, vol. 13, no. 4, s. 441–459.

Di Gregorio Silvana (2000) *Using NVivo for Your Literature Review* [dostęp 18 października 2013 r.]. Dostępny w Internecie: http://www.sdgassociates.com/downloads/literature_review.pdf.

Evers Jeanine C. (2011) *From the Past into the Future. How Technological Developments Change Our Ways of Data Collection, Transcription and Analysis*. „Forum Qualitative Social Research”, vol. 12, no. 1 [dostęp 14 października 2013 r.]. Dostępny w Interne-

cie: <http://www.qualitative-research.net/index.php/fqs/article/view/1636>.

Glaser Barney G., Strauss Anselm L. (1999) *The Discovery of Grounded Theory. Strategies for Qualitative Research*. New Brunswick, NJ: Aldine Transaction.

Głowacki Kamil (2011) *Półmetek perspektywy 2007–2013. Jak wykorzystać pozostałe środki*. Fundacja Studiów Europejskich – Instytut Europejski [dostęp 2 października 2013 r.]. Dostępny w Internecie: http://WWW.ie-ries.com.pl/archiwum/artykuly/RIES_201110141304_Polmetek_II.pdf.

Hutchison Andrew John, Johnston Lynne Halley, Breckon Jeff David (2010) *Using QSR-NVivo to Facilitate the Development of a Grounded Theory Project: An Account of a Worked Example*. „International Journal of Social Research Methodology”, vol. 13, no. 4, s. 283–302.

Kolasa Władysław Marek (2000) *Kartka z dziejów komputeryzacji bibliotek polskich*. „Bibliotekarz”, nr 3 s. 7–11.

Konecki Krzysztof (2000) *Studia z metodologii badań jakościowych. Teoria ugruntowana*. Warszawa: Wydawnictwo Naukowe PWN.

----- (2012) *Wizualna teoria ugruntowana. Podstawowe zasady i procedury*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 12–45 [dostęp 09 października 2013 r.]. Dostępny w Internecie: <http://www.qualitativesociologyreview.org>.

Kuhn Thomas S. (2001) *Struktura rewolucji naukowych*. Przełożyła Helena Ostromęcka. Warszawa: Fundacja Aletheia.

Miller Piotr (2012) *Wprowadzenie do obserwacji online: warianty i ograniczenia techniki badawczej*. „Przegląd Socjologii Jakości-

owej”, t. 8, nr 1, s. 76–97 [dostęp 11 października 2013 r.]. Dostępny w Internecie: <http://przegladsocjologiijakosciowej.org>.

Niedbalski Jakub (2013) *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 126–165 [dostęp 11 października 2013 r.]. Dostępny w Internecie: <http://przegladsocjologiijakosciowej.org>.

Parfieniuk Marek (2012) *Zarządzanie danymi bibliograficznymi za pomocą narzędzi informatycznych* [w:] Bazyl Poskrobki, red., *Metody badań naukowych z przykładami ich zastosowania*. Białystok: Wydawnictwo Ekonomia i Środowisko, s. 217–237 [dostęp 20 października 2013 r.]. Dostępny w Internecie: http://www.bir-gow.pl/index.php?option=com_remository&Itemid=13&func=startdown&id=103.

Practice Standard for Work Breakdown Structures. Second Edition (2006) Newtown Square, PA: Project Management Institute [dostęp 5 maja 2013 r.]. Dostępny w Internecie: <http://www.pmi.org>

Cytowanie

Głowacki Kamil (2014) *Organizacja procesu badawczego a oprogramowanie do organizacji wiedzy i zarządzania projektem badawczym*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 162–195 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: www.przegladsocjologiijakosciowej.org.

Organization of the Research Process in the Context of Knowledge Organization Software and Research Project Management Software

Abstract: The article introduces the reader to the area of computer programs that can contribute to work on a research project in each of its phases and on the organizational and substantive level. The aim of the article is not the comparison and presentation of all of the programs, but the presentation of individual representatives of programs that can be compiled into a single functional system. A system which facilitates such operations like project management, knowledge and bibliographic data organization, and analysis of empirical material with the use of CAQDAS (Computer Assisted Qualitative Data Analysis Software).

Keywords: CAQDAS, organization and knowledge management, bibliographical data management, research facilitating software, Grounded Theory Methodology

[haanagroup.ir/file/262f6c76cfd3353ea116233443052880e0fcfe5f.pdf](http://www.haanagroup.ir/file/262f6c76cfd3353ea116233443052880e0fcfe5f.pdf).

Schönfelder Walter (2011) *CAQDAS and Qualitative Syllogism logic—NVivo 8 and MAXQDA 10 Compared*. „Forum Qualitative Social Research”, vol. 12, no. 1 [dostęp 21 października 2013 r.]. Dostępny w Internecie: <http://WWW.qualitative-research.net/index.php/fqs/article/viewArticle/1514>.

Schwalbe Kathy (2011) *Information Technology Project Management*. Boston, MA: Course Technology.

Silverman David (2009) *Prowadzenie badań jakościowych*. Przełożyły Małgorzata Głowacka-Grajper, Joanna Ostrowska. Warszawa: Wydawnictwo Naukowe PWN.

Smyth Robyn (2008) *Exploring Congruence between Habermasian Philosophy, Mixed-method Research, and Managing Data Using NVivo*. „International Journal of Qualitative Methods”, vol. 5, no. 2, s. 131–145 [dostęp 20 października 2013 r.]. Dostępny w Internecie: <http://wigan-ojs.library.ualberta.ca/index.php/IJQM/article/view/4395>.

Wilk Katarzyna M. (2001) *Komputerowe wspomaganie jakościowej analizy danych*. „ASK. Research & Methods”, nr 10, s. 49–63.

Artur Piszek**Krzysztof Stachura**
Uniwersytet GdańskiDOI: <https://doi.org/10.18778/1733-8069.10.2.10>**Evernote: zastosowanie notatnika internetowego do badań jakościowych**

Abstrakt W artykule przedstawiono aplikację Evernote jako prostą i darmową alternatywę dla profesjonalnych pakietów przeznaczonych do jakościowej analizy danych, przy jednoczesnym wskazaniu ograniczeń opisywanego narzędzia. Wymieniono funkcje przydatne badaczowi do zastosowań analitycznych oraz wskazano techniki pozwalające na aplikację metodologii teorii ugruntowanej przy pomocy tego programu. Wskazano także podobieństwa i różnice pomiędzy Evernote a pakietem NVivo 10 oraz nakreślono nowe perspektywy analityczne na najbliższą przyszłość związane z rozwojem Evernote.

Słowa kluczowe Evernote, CAQDA, badania jakościowe, organizowanie informacji, NVivo

Artur Piszek, absolwent informatyki na Uniwersytecie Warszawskim oraz psychologii na Akademii Pedagogiki Specjalnej, pracuje jako inżynier w Samsung Electronics Polska. Interesuje się szczególnie wspomaganą komputerowo behawioralną zmianą zachowań, zagadnieniami związanymi z ludzką pamięcią oraz wpływem Internetu na przemiany społeczne.

Adres kontaktowy:e-mail: artur@piszek.com<http://piszek.com>

Krzysztof Stachura, dr socjologii, pracuje w Instytucie Filozofii, Socjologii i Dziennikarstwa UG. Członek Zarządu Sekcji Socjologii Komunikacji, Wiedzy i Kultury Międzynarodowego Towarzystwa Socjologicznego. Prezes Fundacji Ośrodek Badań i Analiz Społecznych. Do jego głównych zainteresowań badawczych należą socjologią nowych mediów i metodologią badań online.

Adres kontaktowy:

Instytut Filozofii, Socjologii i Dziennikarstwa

Uniwersytet Gdański, ul. Bażyńskiego 4, 80-952 Gdańsk

e-mail: kastachura@gmail.com**Powszechność metod analizy jakościowej w programach komputerowych**

Pamięć komputerowa znacznie lepiej nadaje się do przechowywania, indeksowania i wyszukiwania fragmentów tekstu niż ludzki umysł czy zapiski na papierze. Dzięki nieograniczonej cierpliwości komputer potrafi z łatwością przeszukiwać duże zbiory informacji, tworzyć połączenia między nimi i zastępować godziny żmudnej pracy jednym kliknięciem myszy. Jak twierdzi Brosz, „potencjał ten należy w sposób kontrolowany i świadomy wykorzystywać z jednej strony w celu podniesienia jakości opracowań, a z drugiej – aby rozsądnie dysponować czasem” (2012: 101).

Strategia indeksowania, porównywania i generalizacji (strategia *code and retrieve*; Niedbalski, Ślęzak 2012) materiałów tekstowych jest tradycyjnie uważana za część domeny teorii ugruntowanej. Niektórzy badacze, między innymi Kelle (1997), twierdzą, że może ona być z powodzeniem stosowana także w analizie dyskursu czy krytycznej etnografii. Warto jednak zauważyć, że indeksowanie i porównywanie informacji zakodowanych w formie pisemnej stanowi na tyle uniwersalny sposób generalizacji wiedzy, iż leży nie tylko u podwalin całego świata nauki, ale także współczesnego społeczeństwa informacyjnego. Bardziej zaawansowane mechanizmy używane przez badaczy, takie jak tworzenie indeksów czy odsyłacze do fragmentów tekstu, są techniką zapożyczoną chociażby z Pisma Świętego, w którym znając numer wersetu, czytelnik może z łatwością odszukać potrzebną informację (Kelle 1997).

Wobec przydatności tych i innych strategii obchodzenia się ze zbiorami wiedzy, producenci opro-

gramowania przekształcili je w algorytmy komputerowe, w wyniku czego obecnie „programy komputerowe pomagają usystematyzować postępowanie badawcze, które od bardzo dawna stanowi podstawę analizy tekstów, na które składa się indeksowanie oraz wstawianie wewnętrznych odnośników” (Bieliński, Iwińska, Rosińska-Kordasiewicz 2007: 91). Jednak wykorzystanie tych technik nie ogranicza się jedynie do aplikacji wspomagających pracę badacza. System dokumentów tekstowych powiązanych referencjami (język HTML i protokół HTTP) stanowi podstawę sieci stron WWW (Fielding i in. 1999), a idea kategoryzowania fragmentów tekstu poprzez kodowanie jest znakomicie widoczna nie tylko w wewnętrznych systemach działania baz danych, ale także w postaci na przykład tak zwanych *hashtagów* serwisu Twitter, pozwalających na odnalezienie wszystkich wiadomości oznaczonych pewnym kodem alfanumerycznym zaczynającym się od znaku #. Obecnie, w erze dominacji mediów elektronicznych, niemal wszyscy spotykamy się z dużą ilością informacji tekstowej. Listy zakupów, zadania do wykonania w pracy, szczegóły zakupionego przez Internet biletu lotniczego, instrukcje dojazdu do domku letniskowego, notatki ze spotkania biznesowego, etykieta tego znakomitego wina czy ciekawy artykuł znaleziony w sieci, odłożony do przeczytania na później to zbiory treści, które można usystematyzować. Takiego zadania podjął się Phil Libin, twórca programu Evernote.

Możliwości Evernote

Evernote został „zaprojektowany tak, aby ułatwić przeszukiwanie i zbieranie informacji” (Evernote

Corporation 2013) i jako elektroniczny notatnik posiada szerokie możliwości gromadzenia, porządkowania i strukturyzowania danych, co stanowi znaczną część pracy badacza w podejściu jakościowym (Seale 2008). Aplikacja ta bardziej spełnia kryteria usługi niż programu. W domyślnej konfiguracji, dane użytkownika znajdują się na centralnym serwerze (tzw. „chmurze”) i użytkownik uzyskuje do nich dostęp przez specjalną stronę internetową, a także za pomocą dedykowanych aplikacji na komputery z systemami Windows i OSX¹, tablety i telefony komórkowe z systemami Android oraz iPhone i iPad. Aplikacje komunikują się z centralnym serwerem za pośrednictwem synchronizacji, dzięki czemu mogą działać bez podłączenia do Internetu, a wymieniać informacje z centralnym serwerem, gdy takie połączenie będzie dostępne.

Dzięki ulokowaniu danych w centralnym repozytorium badacz może udostępnić całe zbiory współpracownikom, przyspieszając w ten sposób wymianę informacji oraz unikając uciążliwego i często prowadzącego do błędów proceduru wymieniania zbiorów dokumentów kanałami takimi jak e-mail.

Pewne wątpliwości niektórych badaczy może budzić fakt przechowywania kopii notatek na zewnętrznym serwerze. Producent zapewnia, że dane te są zaszyfrowane i nikt niepowołany nie ma do nich dostępu. Gdyby jednak użytkownik chciał zrezygnować z synchronizacji, może utworzyć nowy notatnik i wybrać opcję „notatnik lokalny”. W ten sposób notatki, które się w nim znajdują, nie będą

przesyłane na serwer. Nie będą do nich jednak mieli dostępu inni badacze ani oprogramowanie rozpoznające tekst w obrazach. Notatki takie nie pojawią się też na żadnym innym urządzeniu.

Evernote nie jest jedynym programem do tworzenia elektronicznych notatek. Oprócz niego istnieją także między innymi Microsoft OneNote, Simplenote, Springpad i inne, a funkcję tę mogą pełnić także tradycyjne edytory tekstu. Evernote cechuje się jednak szeregiem funkcjonalności, które go wyróżniają. Pozwala między innymi na:

- organizację notatek poprzez przypisywanie wielu znaczników (kodów) zorganizowanych w strukturę drzewa,
- działa na komputerach z systemem OSX i Windows, tabletach i telefonach komórkowych,
- posiada narzędzia wspierające kopiowanie i zachowywanie informacji ze stron WWW oraz dokumentów tekstowych,
- sprawnie przeszukuje całą bazę zgromadzonych danych, łącznie z zawartością plików PDF i dokumentów pakietu MS Office oraz tekstem widocznym na obrazkach znajdujących się wewnątrz notatek.

Praca z programem rozpoczyna się od założenia konta na witrynie Evernote.com. Evernote posiada trzy wersje: darmową², *premium* i *business*. Różni-

² Więcej o możliwościach bezpłatnego oprogramowania wspomagającego analizę danych jakościowych w: Niedbalski (2013).

¹ OSX jest systemem operacyjnym używanym przez firmę Apple w komputerach z serii Mac, Mac Pro, MacBook i MacBook Pro.

ce pomiędzy pierwszymi dwiema przedstawiono w tabeli 1, zaś wersja *business* wychodzi poza obszar zainteresowań niniejszego artykułu. Jak już wspomniano, Evernote jest bardziej usługą niż programem i w konsekwencji nie oferuje możliwości wykupienia licencji bezterminowej. Opłacając wersję *premium*, badacz zyskuje lepsze możliwości kolaboracji, przeszukiwania bazy (wyszukiwanie następuje nie tylko w notatce, obrazku, ale także

w treści załączonych plików MS Office i PDF) oraz wyższy miesięczny limit nowych danych (oznacza to, że w wersji darmowej można przesłać 60 MB nowych danych miesięcznie, jednak całkowita wielkość bazy nie jest ograniczona). Nic nie stoi na przeszkodzie, by badacz wykupił licencję *premium* na czas trwania projektu, a po jego zakończeniu powrócił do wersji darmowej, utrzymując dostęp do swoich danych.

Tabela 1. Różnice między wersją darmową i *premium*.

FUNKCJONALNOŚĆ	KONTO DARMOWE	KONTO PREMIUM
Miesięczny limit przesyłu danych na serwery Evernote	60 MB	1 GB
Rozmiar jednej notatki (z załącznikami)	50 MB	100 MB
Możliwość tworzenia notatników <i>offline</i> na urządzeniach mobilnych	NIE	TAK
Udostępnianie notatnika współpracownikom	Tylko do odczytu	Do odczytu i edycji
Wyszukiwanie w treści PDF i dokumentów MS Office	NIE	TAK

Źródło: opracowanie własne.

Po założeniu konta należy pobrać plik z programem i go zainstalować. Przy pierwszym uruchomieniu wpisuje się dane do założonego wcześniej konta i można rozpocząć pracę. Rodzimy użytkownik może szczególnie docenić polską wersję językową, której często brak w większości tego typu programów, zwłaszcza w wersji bezpłatnej.

Aby zilustrować praktyczne możliwości programu, przygotowano następujące hipotetyczne scenariusze badań:

1. Za pomocą aplikacji Evernote na smartfony badacze mogą katalogować miejsca użytku publicznego. Wykonując na urządzeniu notatkę ze zdjęciem, aplikacja automatycznie zapisuje czas i lokalizację wykonania fotografii. Ponadto za pomocą znaczników badacze mogą od razu dodawać odpowiednie kategorie kodowe, takie jak „funkcja_socjalizacyjna”, „funkcja_rekreacyjna” i tym podobne. Tymczasem ich współpracownicy pracujący przy komputerach będą na żywo obserwować rosnącą bazę miejsc do dalszej analizy, a na widoku mapy naturalne skupiska miejsc

o określonej funkcji lub profilu powiążą ze specyfiką dzielnicy lub obszaru. Możliwość geograficznej lokalizacji notatki może być szczególnie ciekawa dla socjologów czy geografów społecznych studiujących przemiany w przestrzeni publicznej miast. Miejsca ulubione przez artystów ulicznych, trasy wędrowki wolontariuszy WOŚP czy lokalizacja części metropolii uznawanych za niebezpieczne to dane, których przedstawienie na mapie może okazać się ciekawe i wartościowe na dalszych etapach analizy.

2. Analizując rodzaje wypowiedzi fanów zespołów muzycznych, badacze mogą użyć wtyczki Evernote Web Clipper, by łatwo skopiować krótkie komunikaty badanych w serwisie Twitter (tzw. *tweety*) skierowane do swoich idoli i później katalogować je, przypisując odpowiednie kody.

3. Przy obserwacji uczestniczącej mającej na celu opis specyfiki rozmów nawiązywanych podczas wędrowek turystów górskich badacz bez dostępu do komputera może posługiwać się aplikacją na telefony i tablety, jednocześnie kodując dane i prowadząc ich wstępną analizę. Podejście takie jest szczególnie pożyteczne w badaniach opartych na teorii ugruntowanej, gdzie iteracyjny cykl zbierania informacji i analizy pozwala na lepsze dopasowanie teorii do studiowanego empirycznie wycinka rzeczywistości.

Reprezentacja informacji

Podstawową jednostką informacji w programie Evernote jest notatka składająca się zwykle ze sformatowanego tekstu z załącznikami. W tekście tym można używać tabel, wstawiać obrazy, zmieniać

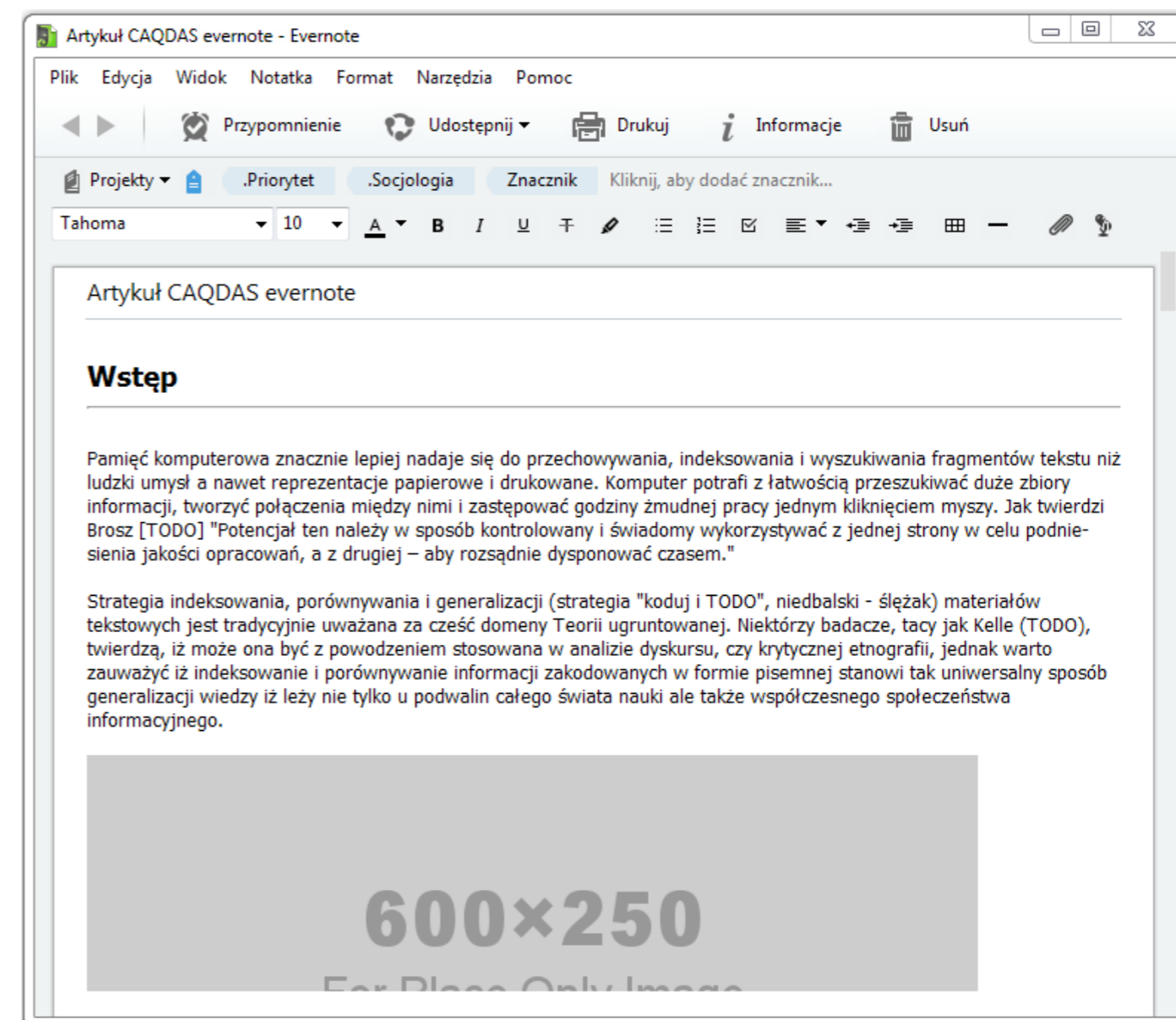
czcionki i używać innych podstawowych narzędzi, takich jak zakreślanie, pogrubianie i tym podobne.

Do każdej notatki można dołączyć załączniki. Nagrania audio, pliki PDF oraz obrazy wyświetlane są bezpośrednio w oknie edycji notatki, co pozwala na posługiwanie się nimi bez konieczności stosowania zewnętrznego oprogramowania. Pliki innego typu, takie jak filmy, dokumenty pakietu MS Office i inne można przechowywać w notatce i otwierać za pomocą stosownych aplikacji. Przykładowe okno edycji notatki przedstawiono na rysunku 1.

Dodatkowo każda notatka zawiera zestaw atrybutów: długość i szerokość geograficzną, adres internetowy źródła danych dla informacji pochodzących z Internetu oraz czas utworzenia i czas ostatniej modyfikacji.

Evernote posiada specjalne udogodnienia do pracy z treścią pochodzącą z serwisów WWW. Za pomocą specjalnej wtyczki do przeglądarki WWW (Evernote Web Clipper), badacz może wyciąć interesujący go fragment witryny z zachowaniem formatowania, tabel czy zdjęć. Często jednak układ treści na stronie WWW jest tak skonstruowany, na przykład poprzez umieszczenie reklam, że badaczowi może sprawiać trudności oddzielenie interesujących go treści od dodatków interfejsu witryny. W takim przypadku Evernote Web Clipper pozwala na przełączenie w tryb *Uproszczony artykuł*, który usuwa z treści większość zbędnych elementów oraz łączy wielostronicowe teksty w jeden ciąg. Evernote Web Clipper posiada także specjalne narzędzia do opisywania wycinanej treści, takie jak strzałki, zakreślanie oraz *pikselator* – specjalny

Rys. 1. Okno edycji notatki



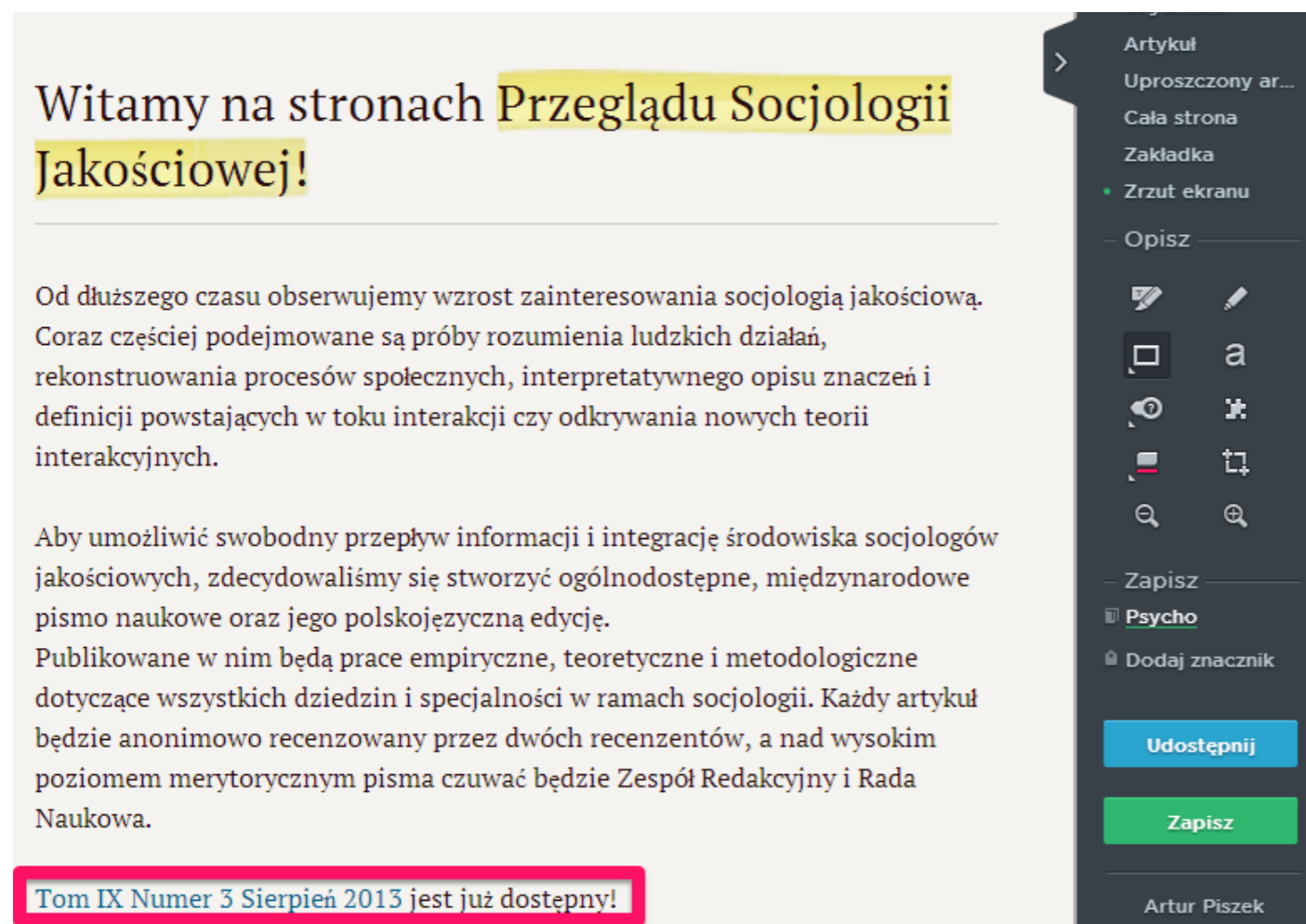
Źródło: opracowanie własne.

stempel pozwalający na przykład na ukrycie tożsamości osób znajdujących się na zdjęciach. Przykład działania Evernote Web Clipper przy współpracy z witryną „Przeglądu Socjologii Jakościowej” przedstawiono na rysunku 2.

W pracy badacza Evernote Web Clipper może być szczególnie użyteczny przy badaniach netnogra-

ficznych blogów, for internetowych i serwisów społecznościowych. Pozwala na wycięcie konkretnego fragmentu witryny do dalszej analizy z zachowaniem odnośnika do oryginału. Takie kopiowanie treści jest dobrą praktyką ze względu na ryzyko usunięcia źródłowej witryny przez administratora lub autora i utracenie do niej dostępu przez badacza.

Rys. 2. Przykład działania Evernote Web Clipper



Źródło: opracowanie własne.

Aplikacje Evernote na tablety i telefony posiadają możliwość nagrywania dźwięku oraz przechwytywania zdjęć za pomocą aparatu danego urządzenia. Funkcjonalności te pozwalają na prowadzenie nagrania czy tworzenie dokumentacji fotograficznej niezbędnej w niektórych typach obserwacji.

Sposób organizacji informacji

Podobnie jak w metodologii teorii ugruntowanej (Niedbalski, Ślęzak 2012), a także innych podejściach badawczych, do organizacji informacji wyko-

rzystano kody, które zaimplementowano za pomocą dwóch struktur: notatników oraz znaczników.

Notatniki działają podobnie do folderów na komputerze. Każda notatka może się znajdować tylko w jednym notatniku. Warto przy tym zauważyć, że poszczególnych notatników nie można osadzać wewnątrz innych, toteż nie nadają się najlepiej na reprezentację kodów w myśl metodologii teorii ugruntowanej. Tę funkcję znacznie lepiej pełnią znaczniki, które można osadzać wewnątrz innych, tworząc tym samym hierarchiczną strukturę pomocną przy pro-

Rys. 3. Interfejs Evernote na Apple iPad



Źródło: opracowanie własne.

cesie budowania teorii (Glaser, Strauss 1967; Konecki 2000). Do każdej notatki można dodać dowolną ich liczbę. Poszczególne wartości atrybutów można także przedstawić za pomocą znaczników, jednak taka operacja pokaże je na skali nominalnej. Przykładowo, dla atrybutu *pleć* badacz może utworzyć znaczniki *plećM* i *plećK*, a dekadę urodzenia poprzez *generacjaX*, *generacjaY*, *generacjaZ*. Takie właśnie kodowanie atrybutów przedstawiono na rysunku 4.

Znaczniki mogą być zorganizowane w hierarchiczną strukturę drzewa. Oznacza to, że można zagłębić

jeden kod w drugim i w ten sposób uzyskać związki rodzic–dziecko, a rozwijając przedstawiony wcześniej przykład, utworzyć kod *pokolenie*, a w nim zagłębić kategorie *generacjaX*, *generacjaY* i *generacjaZ*. Nie jest możliwe tworzenie innego typu powiązań przypominających strukturę grafu i łączących dowolne kategorie kodowe, tak jak ma to miejsce na przykład w NVivo. W celu utworzenia map koncepcyjnych badacz będzie musiał posłużyć się innym oprogramowaniem. Darmowymi, godnymi polecenia opcjami są XMind (Xmind Ltd. 2013) i MindMeister (MeisterLabs 2014).

Znacznik jest przypisywany do notatki poprzez wpisanie jego nazwy w pole „Kliknij, by dodać znacznik”. Jeśli został już utworzony, pojawi się po wpisaniu początku nazwy na rozwijanej liście. W przeciwnym razie, program jednocześnie utworzy go i przypisze do aktualnej notatki. Nowe znaczniki badacz podaje po przecinku, przez co przypisywanie wielu kodów do jednej notatki jest sprawniejsze. Drugą metodą kodowania jest przeciągnięcie notatki na nazwę danego znacznika w hierarchicznym drzewie. Proces umieszczania notatki w notatniku przebiega podobnie – poprzez przeciągnięcie lub kliknięcie nazwy obecnego notatnika na górnym pasku notatki.

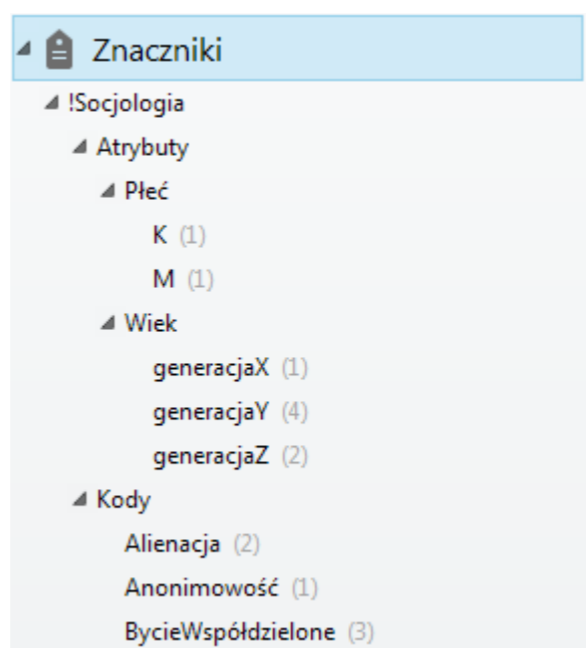
Na drzewie notatników i znaczników w nawiasach program wyświetla liczbę notatek należących do danej taksonomii. W ten sposób jednym rzutem oka badacz może zauważyć, które kategorie kodowe są częściej wykorzystywane, i w trakcie prac badawczych dostosować dalsze kroki zgodnie z założeniami teorii ugruntowanej.

Warto nadmienić, że zastosowanie Evernote jako narzędzia do kodowania danych z wywiadu może narażać na pewnych trudności. Brakuje w nim na przykład funkcji, która pozwalałaby na przypisywanie kodów poszczególnym fragmentom tekstu lub wierszom, tak jak jest to możliwe w NVivo (Niedbalski, Ślęzak 2012) czy OpenCode (Niedbalski 2012). Okno edycji tekstu wykorzystane w programie pozbawione jest możliwości tworzenia komentarzy i przypisów, a znaczniki reprezentujące kody mogą być przypisane do całej notatki, a nie konkretnego miejsca w jej treści. Istnieją naturalnie sposoby obejścia tej limitacji, takie jak tworzenie oddzielnej

notatki na każdy paragraf, jednak podejmowanie takich procedur może bardziej badaczowi pracę utrudniać niż w niej pomagać.

Na uwagę zasługuje także możliwość tworzenia odnośników pomiędzy notatkami. Gdy badacz kliknie prawym przyciskiem na notatce i wybierze „Kopiuj łącze notatki”, a następnie wklei takie łącze w innej notatce, tworzy tym samym powiązanie pomiędzy tymi dwiema informacjami. Łączy takie tworzą ze zbioru notatek *Hipertekst* i pozwalają na nielinowe poruszanie się po treści i, jak zauważyli Coffey, Holbrook i Atkinson (1996), znacznie bardziej interaktywną relację pomiędzy tekstem a czytelnikami. Łączy takie są także istotną pomocą dla badacza w tym sensie, że można jednym kliknięciem zweryfikować lub przypomnieć sobie informację kryjącą się pod odnośnikiem.

Rys. 4. Przykład hierarchicznej struktury kodów utworzonej za pomocą znaczników



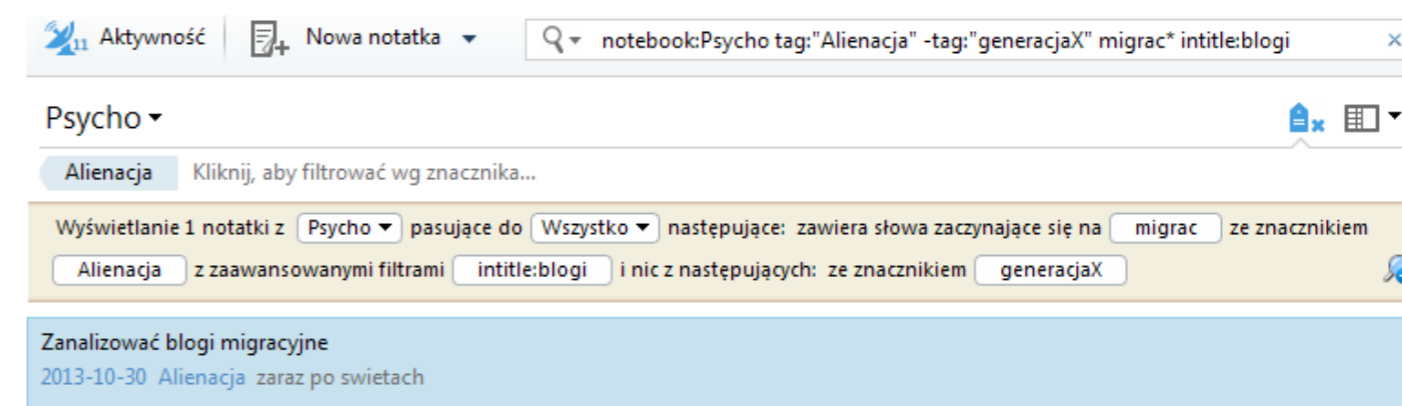
Źródło: opracowanie własne.

Po kliknięciu danego znacznika lub notatnika użytkownik zobaczy listę notatek opisanych przez wybraną taksonomię. Często taka funkcjonalność jest niewystarczająca – z tego powodu twórcy wyposażyli to oprogramowanie w zaawansowany system wyszukiwania zapisanych informacji.

Wpisanie hasła w pasek wyszukiwania przeszkuduje jednocześnie notatki, załączone do nich dokumenty PDF, pliki pakietu MS Office oraz znajdujące się w notatkach obrazy. Ta ostatnia funkcjonalność jest dostępna po synchronizacji. Gdy obrazek znajdzie się już na centralnym serwerze, obecne tam oprogramowanie rozpoznaje na nim tekst i koduje transkrypcję w notatce. Niestety użytkownik nie ma do niej dostępu (co bez wątpienia przydałoby się wielu badaczom), jednak możliwość przeszukiwania zeskanowanych lub sfotografowanych dokumentów bez uprzedniego przepisywania jest warta odnotowania.

Wyniki wyszukiwania można zawęzić do znacznika lub notatnika poprzez uprzednie kliknięcie nazwy danej taksonomii i następnie wpisanie hasła w pasek wyszukiwania. Możliwe też jest stosowanie skomplikowanych operatorów wyszukiwania, takich jak wyszukiwanie w konkretnym znaczniku lub notatniku, wyszukiwanie notatek z konkretnym tytułem, utworzonych bądź zmodyfikowanych w określonym czasie, z pewnymi rodzajami załączników oraz utworzonych w danej lokalizacji geograficznej. Operatory wyszukiwania można łączyć i przykładowo hasło *notebook:Psycho tag:"Alienacja" -tag:"generacjaX" migrac* intitle:blogi* przedstawi notatki z notatnika *Psycho*, opatrzone kodem *Alienacja*, dotyczące osób badanych spoza pokolenia X, zawierające w tytule *blogi*, a w treści notatki słowo zaczynające się na *migrac* (czyli migracje, migracyjnych itd.). Stosowny zrzut ekranu przedstawiający interpretację zapytania wyszukiwania przez Evernote zamieszczono na rysunku 5.

Rys. 5. Interpretacja przez program wyszukiwania *notebook:Psycho tag:"Alienacja" -tag:"generacjaX" migrac* intitle:blogi*



Źródło: opracowanie własne.

Badacz często używający konkretnej formuły może ją zapisać w panelu podręcznym dla późniejszych

zastosowań poprzez kliknięcie przycisku „Zapisz wyszukiwanie...”.

Analiza jakościowa w Evernote

Evernote nie został stworzony z myślą o analizie jakościowej. Aplikacja powstała jako narzędzie do sprawnego kopiowania, archiwizowania i przeszukiwania zbiorów informacji. Jednak według Kellego (1997) także typowe programy z rodziny CAQDA (Computer-Assisted Qualitative Data Analysis) należą do tej samej kategorii narzędzi służących do administracji i archiwizacji danych. Są one w grun-

cie rzeczy programami do automatyzowania żmudnych czynności porządkowania i archiwizowania tekstów, wykonywanych w naukach hermeneutycznych od stuleci. Dla przybliżenia jego możliwości jako aplikacji do analizy jakościowej w tabeli 2 przedstawiono porównanie funkcjonalności Evernote oraz aplikacji specjalnie zaprojektowanej do wykorzystania przez badaczy jakościowych (Niedbalski, Ślęzak 2012), a mianowicie pakietu NVivo w wersji 10.

Tabela 2. Porównanie zastosowań NVivo i Evernote do analizy jakościowej.

FUNKCJONALNOŚĆ	NVIVO	EVERNOTE
Tworzenie notatek	Tworzenie <i>memos</i> za pomocą edytora tekstu	Tworzenie notatek za pomocą edytora tekstu
Kategoryzowanie treści	Za pomocą węzłów i atrybutów	Za pomocą znaczników
Wewnętrzne łączenie treści	Za pomocą „See also links”	Za pomocą łączy notatek
Wstawianie plików większych niż 100 MB	TAK	NIE
Zaawansowana edycja kodów	Tworzenie związków między kodami	NIE
Współpraca z innymi badaczami w czasie rzeczywistym	Za pomocą NVivo Server	Za pomocą notatników współdzielonych
Wspierane systemy operacyjne	Windows	Windows, OSX, Android, Apple iPad i iPhone
Upublicznianie informacji	Poprzez „Ministry”	Poprzez publiczne notatniki w sieci WWW
Gromadzenie informacji z Internetu	NVivo NCapture	Evernote Web Clipper
Praca w terenie	NIE	Poprzez telefon lub tablet
Rozszerzanie za pomocą aplikacji	NIE	Poprzez otwarte API
Komentowanie i kategoryzowanie fragmentów tekstu	Poprzez węzły i <i>memo</i>	NIE
Przechowywanie treści multimedialnych	TAK	TAK
Komentowanie i kategoryzowanie treści multimedialnych	TAK	NIE
Analiza numeryczna treści	Poprzez raporty, zapytania, chmurę tagów itp.	NIE
Analiza graficzna treści	Diagramy i wykresy	NIE

Źródło: opracowanie własne.

Podobnie jak NVivo, Evernote posiada możliwość tworzenia notatek w edytorze tekstu, kategoryzowania ich, tworzenia wewnętrznej sieci powiązań oraz wstawiania treści multimedialnych. W przypadku Evernote w wersji *premium* treść notatki nie może przekraczać 100 MB, co może być kłopotliwe, gdy analizie poddaje się materiały audiowizualne. Oba programy pozwalają na współpracę z innymi badaczami w czasie rzeczywistym. W przypadku NVivo należy w tym celu zakupić oddzielne oprogramowanie (NVivo server), a także zapewnić komputer (serwer), na którym będzie ono cały czas dostępne. W przypadku Evernote można skorzystać z funkcji notatników współdzielonych. Dzięki synchronizacji Evernote wspiera większą liczbę aplikacji. Dostępne są wersje na tablety, telefony, komputery firmy Apple oraz każdy inny system dzięki przeglądarce internetowej. Dzięki otwartemu API (*Application Interface*) zewnętrzni producenci oprogramowania mogą tworzyć dodatkowe aplikacje współpracujące z Evernote. Przykładami takich aplikacji są na przykład Skitch do adnotacji obrazów, tworzenia podkreśleń i komentarzy na zdjęciach oraz KustomNote (Kustomnote.com), pozwalający na kreowanie internetowych ankiet i formularzy, których wyniki pojawiają się bezpośrednio na koncie Evernote. Szczególnie ta ostatnia propozycja wydaje się interesująca dla badaczy.

NVivo, w przeciwieństwie do Evernote, posiada możliwość tworzenia zaawansowanych związków między węzłami, komentowanie i kategoryzację treści multimedialnych oraz numeryczną i graficzną analizę treści oraz autokodowanie. Nie ulega wątpliwości, że NVivo jest potężnym narzędziem specjalnie zaprojektowanym do analizy jakościowej i wieloma funkcjonalnościami bezapelacyjnie

przerasta Evernote w tym zastosowaniu. Jednak interfejs Evernote jest bardziej intuicyjny i wygodniejszy, a możliwość korzystania z niego na różnorodnych urządzeniach, włączając terminale mobilne, takie jak tablet czy telefon, pozwala na sprawniejsze gromadzenie zbiorów danych do dalszej analizy. Efektywne korzystanie z pakietu NVivo wymaga nabycia określonego poziomu biegłości, przez co początkujący użytkownicy mogą się czuć zagubieni. Ponadto Evernote posiada wersję darmową, z której można korzystać dowolnie długo i, co szczególnie istotne, bez ryzyka utraty dostępu do zgromadzonych danych, jak ma to miejsce w przypadku 30-dniowej wersji testowej NVivo.

Poprzez bardziej uniwersalne przeznaczenie Evernote dobrze się sprawdza w zastosowaniach „okołoprojektowych”. Przykładowo, wielu badaczy, wliczając w to autorów niniejszego artykułu, zapoznając się z literaturą naukową, gromadzi cytaty i potrzebne fragmenty z publikacji drukowanych poprzez fotografowanie ich wybranych ustępów za pomocą telefonu komórkowego. Tym sposobem, dzięki automatycznej synchronizacji oraz rozpoznawaniu tekstu przez serwery Evernote, cytat, z którym badacz zapoznawał się w bibliotece, może zostać z łatwością odszukany na własnym komputerze, po uprzednim wpisaniu jego fragmentu w pole wyszukiwania.

Evernote pełni zatem rolę menadżera tekstów, opisanego przez Bielińskiego, Iwińską i Rosińską-Kordasiewicz (2007) jako

program służący do zarządzania bazą tekstów z różnych źródeł (w różnych formatach). Możliwość taka stanowi duże ułatwienie dla tych, którzy

zbierają bardzo wiele różnego rodzaju danych (np. przy badaniach monograficznych), ponieważ podstawą pracy w większości programów CAQDA jest wspólna baza danych określonego formatu. (s. 95)

Jednak badacz nie musi wybierać między dwoma rozwiązaniami. W wersji 10, pod naciskiem użytkowników obu programów, producent pakietu NVivo wprowadził możliwość importu danych z Evernote, z zachowaniem całej treści oraz znaczników, które podczas importu zostają przekształcone w węzły (Qsr 2013). Nie ma zatem przeciwwskazań, by wstępnie informacje zbierać w Evernote, a gdy zajdzie potrzeba lub wymagania projektu przerosną możliwości Evernote, zaimportować całość do NVivo i tam kontynuować pracę.

Evernote i nowe perspektywy analityczne

Wprawdzie znaczna część tego tekstu poświęcona jest konkretnym możliwościom, jakie daje Evernote w zakresie pracy z danymi jakościowymi (docelowo przedstawiając to narzędzie jako alternatywę dla rozbudowanego oprogramowania typu CAQDA), celem opisu jest skłonienie czytelnika do generalnej refleksji – czy i w jakim analitycznym kontekście warto jest włączyć nową aplikację do badawczej skrzynki narzędziowej.

Potraktowanie Evernote jako programu, który pełni rolę wsparcia dla badaczy społecznych w procesie pracy z danymi, jest świadectwem tego, jak poszerza się instrumentarium, z którego można korzystać w prowadzeniu różnych projektów (także tych o rozbudowanym zakresie i strukturze). Dla wielu zastosowań badawczych naturalny będzie wybór bardziej

zaawansowanych narzędzi, takich jak choćby NVivo, do którego Evernote jest w tekście porównany. Nie oznacza to jednak, że aplikacje niepomyślane tradycyjnie jako usługi wspierające proces badawczy nie mogą zostać w nim wykorzystane.

Jest to też zasadne o tyle, że wśród badaczy decydujących się na wykorzystywanie w pracy analitycznej możliwości oferowanych przez nowe technologie informacyjno-komunikacyjne wielu jest takich, którzy przywykają do korzystania jednocześnie z kilku narzędzi wspomagających analizę. Poszczególne narzędzia, w tym także i Evernote, mogą pełnić zarówno funkcję badawczych „czarnych skrzynek”, organizując cały warsztat analityczny, jak i być wykorzystywane jako jedna z większej liczby aplikacji.

Evernote jest kolejnym narzędziem, które pomaga lepiej organizować pracę badacza, a dodatkowo jego stosunkowo prosty interfejs pozwala na szybkie opanowanie podstaw korzystania z programu. Jest to istotne między innymi ze względu na możliwość zapraszania do wspólnej pracy nad różnymi projektami innych osób. Evernote dobrze sprawdza się jako usługa organizująca pracę zespołu, znacznie ją też przy tym usprawnia (także w wymiarze czasowym). Dodatkowo, w ocenie przydatności narzędzia warto wziąć pod uwagę fakt jego dostępności na urządzeniach mobilnych, co pozwala na wykonywanie pracy analitycznej także niestacjonarnie.

Podobnie jak w przypadku oprogramowania dedykowanego bezpośrednio do celów analitycznych, przydatność i skala zastosowania Evernote w procesie badawczym zależą będzie od kilku czynników. Z jednej strony chodzi tu o uwarunkowania tech-

niczne, czyli to, jak rozwijana będzie aplikacja, jaka będzie jej kompatybilność z innymi usługami i tak dalej, z drugiej natomiast – czynniki mentalne, czyli skłonność badaczy do wypróbowywania nowego narzędzia i otwartość na poznanie alternatywnego modelu pracy analitycznej. Obserwowanie tej ewolucji pozwoli na wyciągnięcie wniosków o zakresie zmiany formuły warsztatu badawczego w czasach, gdy możliwości pracy badawczej dają coraz większa liczba programów, aplikacji i usług.

Bibliografia

Bieliński Jacek, Iwańska Katarzyna, Rosińska-Kordasiewicz Anna (2007) *Analiza danych jakościowych przy użyciu programów komputerowych*. „ASK”, nr 16, s. 89–114.

Brosz Maciej (2012) *Komputerowe wspomaganie badań jakościowych. Zastosowanie pakietu NVivo w analizie materiałów nieustrukturyzowanych*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 98–125 [dostęp 20 października 2013 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Coffey Amanda, Holbrook Beverley, Atkinson Paul (1996) *Qualitative Data Analysis: Technologies and Representations*. „Sociological Research Online”, no. 1 [dostęp 15 października 2013 r.]. Dostępny w Internecie: <<http://www.socresonline.org.uk/1/1/4.html>>.

Evernote Corporation (2013) *Remember Everything* [dostęp 23 października 2013 r.]. Dostępny w Internecie: <<http://evernote.com>>.

Fielding, Roy i in. (1999) *Hypertext Transfer Protocol -- HTTP/1.1* [dostęp 17 października 2013 r.]. Dostępny w Internecie: <<http://tools.ietf.org/html/rfc2616>>.

* * *

Czytelnika pragnącego poznać przykłady wykorzystania Evernote w świecie nauki zainteresować mogą witryny WWW http://bit.ly/evernote_academia_1, http://bit.ly/evernote_academia_2 i http://bit.ly/evernote_academia_3, na których zamieszczono przykłady i scenariusze wykorzystania programu w pracach naukowych (zarówno z zakresu nauk społecznych, jak i astronomii, archeologii i prawa).

Glaser Barney, Strauss Anselm (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Company.

Kelle Udo (1997) *Theory Building in Qualitative Research and Computer Programs for the Management of Textual Data*. „Sociological Research Online”, vol. 2 [dostęp 11 października 2013 r.]. Dostępny w Internecie: <<http://www.socresonline.org.uk/2/2/1.html>>.

Konecki Krzysztof (2000) *Studia z metodologii badań jakościowych. Teoria ugruntowana*. Warszawa: Wydawnictwo Naukowe PWN.

MeisterLabs (2014) *MapMeister* [dostęp 20 stycznia 2014 r.]. Dostępny w Internecie: <<http://www.mapmeister.com>>.

Niedbalski Jakub (2013) *Odkrywanie CAQDAS. Wybrane bezpłatne programy komputerowe wspomagające analizę danych jakościowych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

----- (2012) *OpenCode – narzędzie wspomagające proces przeszukiwania i kodowania danych tekstowych w badaniach*

jakościowych. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 220–228 [dostęp 4 października 2013 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Niedbalski Jakub, Ślęzak Izabela (2012) *Analiza danych jakościowych przy użyciu programu NVivo a zastosowanie procedur metodologii teorii ugruntowanej*. „Przegląd Socjologii Jakościowej”, t. 8, nr 1, s. 126–165 [dostęp 20 października 2013 r.]. Dostępny w Internecie: <<http://www.przegladsocjologiijakosciowej.org>>.

Qsr (2013) *NVivo 10* [dostęp 10 października 2013 r.]. Dostępny w Internecie: <http://www.qsrinternational.com/products_nvivo_free-trial-software.aspx>.

Seale Clive (2008) *Wykorzystanie komputera w analizie danych jakościowych* [w:] David Silverman, red., *Prowadzenie badań jakościowych*. Warszawa: Wydawnictwo Naukowe PWN, s. 233–256.

Xmind Ltd. (2013) *XMind: The Most Professional Mind Mapping Software* [dostęp 20 stycznia 2014 r.]. Dostępny w Internecie: <<http://www.xmind.net>>.

Cytowanie

Piszek Artur, Stachura Krzysztof (2014) *Evernote: zastosowanie notatnika internetowego do badań jakościowych*. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 196–210 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.

Evernote: Using Online Notes for Performing Qualitative Research

Abstract: The paper is a presentation of potential ways to use Evernote as a simple and cheap alternative to professional CAQDA (Computer-Assisted Qualitative Data Analysis) software packages, along with pointing specific limitations of the tool. The main functions useful in research and for analytic purposes are depicted, as well as techniques that enable the application of grounded theory methodology using Evernote are. Moreover, similarities and differences between Evernote and NVivo 10 are presented, and the analytical perspectives for the near future concerning the evolution of Evernote are drawn.

Keywords: Evernote, CAQDA, qualitative research, organizing data, NVivo

Piotr Siuda

Uniwersytet Kazimierza Wielkiego w Bydgoszczy

DOI: <https://doi.org/10.18778/1733-8069.10.2.11>**Recenzja książki****Magdalena Kamińska (2011) *Niečne memy. Dwanaście wykładów o kulturze Internetu*. Poznań: Galeria Miejska Arsenał**

Książkę Magdaleny Kamińskiej pod tytułem *Niečne memy* czyta się bardzo dobrze. Jej autorka nie tylko jest kompetentna w opisie podjętego tematu, ale również pisze z pasją oraz w sposób, który sprawia, że lektura jest interesująca zarówno

dla humanisty czy reprezentanta nauk społecznych, jak i „przeciętnego” użytkownika elektronicznej pajęczyny. Każdy, kto zaznajomiony jest z Internetem na najbardziej nawet podstawowym poziomie, związanym chociażby z użyciem różnych narzędzi sieciowych, odnajdzie w omawianym dziele wiele znanych mu motywów – przy tym poddane one zostały naukowemu rozłożeniu na czynniki pierwsze i teoretycznie obudowane. Opisane zostały między innymi prawidłowości komunikacji internetowej, fenomeny konfliktów sieciowych, cybercelebrytów, e-tabloidów (rodzących tak zwaną cyberplotkę), religii cyberprzestrzeni, e-handlu (w tym portali aukcyjnych) czy wspólnot wirtualnych. Autorka zajęła się także zjawiskami fanowskiej aktywności w elektronicznej pajęczynie, zachodzącej w niej zmiany płci (czy generalnie – budowaniem tożsamości) oraz przejawami tak zwanego *group mind power* (blogi, cyberoszustwa grupy Anonymous).

W książce opisana jest formująca się na naszych oczach kultura Internetu. Badaczka wyraża wątpliwość, czy da się jednoznacznie rozstrzygnąć,

czy cyberkultura jest odbiciem i częścią kultury *offline*, czy raczej autonomicznym bytem niemającym z *real life* (określenie Kamińskiej) nic (a przynajmniej niewiele) wspólnego (rozd. 11–12, s. 297). Autorka deklaruje niezamykanie się na obydwie opcje, wydaje się jednak, że czyni to nieco sztucznie, ponieważ wiele prezentowanych w dziele fenomenów jest rozpatrywanych w kontekście pozasieciowym. Analizowane są one (fenomeny) przez pryzmat realności dlatego, że najpierw następuje wymienienie teorii odnoszących się do świata *offline*, a później przyłożenie do niego konkretnego zjawiska z elektronicznej pajęczyny. Na przykład w rozdziale traktującym o sieciowych celebrytach (s. 75–99) badaczka zaczyna od zdefiniowania herosów, gwiazd i celebrytów realnych, dalej informuje, czym jest sława, jak się narodził oraz rozwijał przemysł związany z celebrytami, zahacza o teorie szukające źródeł motywacji, dla których ktoś chce zostać gwiazdą. Wylizać można by dość długo – chodzi o to, że dywagacje o e-celebrytach są poprzedzone analizą nieinternetowego oblicza gwiazdorstwa. Podobnie – jak się rzekło – w wypadku innych fenomenów.

W takim podejściu nie byłoby nic złego, a nawet należałoby je zastosować, gdyby Kamińska jasno zadeklarowała, że cyberkulturę trzeba potraktować jako swoiste odbicie *real life*, a nie nową jakość. Autorka tego jednak nie czyni (co nie jest wadą) i dlatego w analizowaniu różnych zjawisk nie powinna stosować prostego przełożenia ze świata *offline*, a jeśli już się na to zdecydowała, wyraźniej zaznaczyć, że to, co istnieje w sieci, może wcale nie wiązać się z obszarem pozainternetowym. Poruszanie się w jego obrębie oburzy naukowców

podkreślających, że w przypadku elektronicznej pajęczyny mamy do czynienia z jakościową zmianą. Jaskrawym jej przejawem ma być fakt, że w badaniach sieci niezasadne jest nawet odnoszenie się do języka spoza Internetu (Tarkowski 2006). Takie skrajne podejście wyklucza zatem użycie pojęć cybercelebryta czy e-celebryta (trzeba wymyślić nowy termin, nieodwołujący się do *real life*). Powyższych uwag nie należy traktować jako zaniegowania drogi obranej przez badaczkę. Konieczne jest uświadomienie sobie, że ścieżka rozważań o sieci jest niezwykle kręta i najeżona pułapkami wynikającymi z nie do końca jasnych powiązań pomiędzy światem *online* i *offline*. Chwała wszystkim tym, którzy podejmują się wędrówki. Kamińska, wyrażając intelektualne rozterki towarzyszące jej w pisaniu, stwierdziła:

[c]zy jednak pojęciem cyberkultury nie szermuje się na wyrost? Czy nie przejdzie ono do historii kulturoznawczo zorientowanego medioznawstwa jako nigdy nie wprowadzona w życie idea, niezaakceptowany model teoretyczny (...)? O wiele łatwiej byłoby dowieść, że poszczególne cybersrodowiska posiadają własne kultury, a zatem że w Sieci istnieje wiele cyberkultur. Każda z wirtualnych grup ulokowanych w określonych punktach cyberprzestrzeni buduje swoją własną historię, kulturową pamięć, zasób eksplicytnej i milczącej wiedzy dotyczącej lokalnie obowiązujących norm. (s. 11)

Autorka zdecydowała się na scharakteryzowanie owych cybersrodowisk i ich kultur, o czym świadczy układ książki, to jest podzielenie jej na dwanaście rozdziałów traktujących o kolejnych

Piotr Siuda, doktor socjologii; interesuje się socjologią kultury, społecznymi aspektami Internetu oraz edukacją medialną. Autor kilku monografii, w tym: *Religia a Internet* (2010), *Kultury prosumpcji* (2012) oraz *Japonizacja* (2014); publikował w wielu periodykach naukowych (między innymi w „European Journal of Cultural Studies”, „Studia Socjologiczne”, „Kultura i Społeczeństwo”, „Kultura i Edukacja”). Koordynator szeregu projektów badawczych, w tym zadania Prosumpcjonizm pop-przemysłów, Dzieci sieci i Dzieci sieci 2.0. Blog i strona domowa: www.piotrsiuda.pl.

Adres kontaktowy:

Katedra Socjologii

Wydział Administracji i Nauk Społecznych UKW

ul. Ogińskiego 16, 85-092 Bydgoszcz

e-mail: piotr.siuda@gmail.com

fenomenach elektronicznej pajęczyny i powiązanych tylko szeroko rozumianym polem badawczym, jakim jest Internet. Po książce można się spodziewać nieco więcej, jeśli spojrzeć na tytuł, w którym wyraźnie zaznaczone jest, że mowa nie o kulturach, lecz o kulturze. Niewątpliwie po odłożeniu lektury na półkę jest się ukontentowanym, tak dobry był opis cybermiejsc, gdzie obowiązują określone normy. Pozostaje jednak lekki niedosyt, że nie została podjęta próba znalezienia punktów stykowych pomiędzy owymi środowiskami i przez to zaprezentowanie kultury sieci jako bytu całościowego. Pojedyncza informacja o tym, że elementy różnych internetowych enklaw „odnaleźć (...) można w praktyce każdej cyberzbiorowości” (s. 12) to trochę za mało.

Mimo wspomnianych nieścisłości, recenzowana książka jest godna polecenia. Badaczka spojrzała na scharakteryzowane zjawiska w sposób wyczerpujący, ale także niezamykający się na różne punkty widzenia. Można powiedzieć, że Kamińskiej udało się „wyłapać” bardzo dużo sieciowych memów, czyli internetowych jednostek przekazu kulturowego (zob. Blackmore 2000). Z powodu jego wszechstronności, dzieło trzeba potraktować jak podręcznik, który posłużyć może akademikom w celach dydaktycznych, na przykład w prowadzeniu kursów socjologii czy psychologii Internetu. Fakt ten lokuje omawianą pozycję niezwykle wysoko na wydawniczym rynku – jest to bodaj pierwsza tego typu publikacja.

Chociaż niektórym czytelnikom nie spodoba się dość niecodzienne potraktowanie materii językowej, przejawiające się słowotwórczą swobodą

autorki, to jednak należy owo szafowanie zaskakującymi w swojej wymowie wyrazami uznać za przejaw błyskotliwości i jeden z elementów dobrze pojmowanego naukowego krytykanctwa. Badaczka nie boi się wyrażać swojego zdania i korygować poglądów innych, a przy tym wyraźnie zaznacza, jakie tezy i spojrzenia *web studies* już się w owej dziedzinie przeterminowały (przykładem są teorie mówiące o intensywnej zmianie tożsamości płciowych [*gender-switching*] w sieci). Uporczywe trwanie przy starych, słynnych niegdyś teoriach jest wadą wielu innych publikacji traktujących o społecznym wymiarze cyberprzestrzeni. Kamińska nie popełnia tego błędu i jest to godne pochwały zwłaszcza dlatego, że Internet to niezwykle dynamicznie zmieniające się środowisko badawcze.

Wspomniane już zostało, że książka podzielona jest na dwanaście rozdziałów dotyczących różnych aspektów elektronicznej pajęczyny. Na zakończenie warto wspomnieć o kilku ciekawych przemysłeniach autorki odnoszących się do niektórych z owych obszarów. Chodzi o kwestie nieoczywiste, których podjęcie trzeba uznać za wartość dodaną dzieła – takich rozważań, stawiających w nowym świetle znane już badaczom zjawiska, znaleźć można w *Niecnym memach* bardzo wiele. Na przykład, analizując konflikty, walkę i niewidzialną przemoc w Internecie, Kamińska zauważyła, że można owe procesy traktować jako ważne w kontekście podtrzymywania hierarchii między internautami (s. 28). Chodzi o to, że ten, kto potrafi wywołać konflikt oraz orientuje się w zaistniałej sytuacji, to znaczy zna odpowiedni język, a także zachowuje się zgodnie z obowiązującymi normami, jest na drabinie sieciowego statusu wyżej.

Inny ciekawy moment następuje, kiedy autorka przeciwstawia się dominującym obecnie poglądom, jakoby każdy internauta był aktywnym twórcą kultury (przyczyniając się do zniesienia podziału na dwie – nierozdzielne w przypadku innych mediów – sfery, to jest konsumpcji i produkcji) (s. 63). Mimo że istnieją tacy użytkownicy, którym się owo miano należy, większość to po prostu „transmitterzy” treści kulturowych, niekonstruujący nic nowego, a jedynie przekazujący dalej to, co już istnieje.

Przedstawiając fenomen autoprezentacji w elektronicznej pajęczynie, badaczka słusznie stwierdziła rzecz do tej pory niezdiagnozowaną – ukazujący się w sieci internauci poddani są dwóm, paradoksalnie przeciwstawnym, wymogom: z jednej strony wskazane jest, aby ujawnili jak najwięcej (choćby bloger chcący odnieść sukces musi informować czytelników o prywatnych sferach swojego życia), a z drugiej – muszą być świadomi niebezpieczeństw związanych z nadmiernym internetowym ekshibicjonizmem (inni użytkownicy mogą zniekształcić wizerunek danej osoby) (s. 131). Za niezwykle interesujące uznać należy czynione kilka razy uwagi, że w sieci niekoniecznie liczyć się musi anonimowość – wiele środowisk stawia raczej na modernistycznie rozumianą stabilność tożsamości, a nie postmodernistyczną płynność zmiennych identyfikacji. Owa tendencja promowania tego, co stałe, a nie zmienne szczególnie mocno zaznacza się tam, gdzie liczy się płeć internautów (na przykład serwisy randkowe [s. 154]), a także na portalach aukcyjnych, które stabilnych tożsamości potrzebują dla prawidłowego funkcyj-

nowania. Rozpoznawalność sprzedawców jest w końcu warunkiem koniecznym obdarzenia ich zaufaniem (s. 249).

Nieco słabszy moment w rozważaniach następuje wtedy, gdy Kamińska zabiera się za wirtualną wspólnotowość. Szczególnie razi terminologiczny miks niewnoszący zbyt wiele w zrozumienie opisywanych zjawisk. Autorka oddziela wspólnotę wirtualną od wspólnoty *online* (s. 278), nie do końca jednak wiadomo, co badaczka chce osiągnąć, bowiem definiując obydwie formy, podaje te same ich wyróżniki. Wątpliwości pojawiają się również przy dywagacjach o tak zwanych religiach cyberprzestrzeni (Siuda 2010), czyli internetowych humorystycznych miejscach religijnych (s. 207). Kamińska spostrzegła, że wielu akademików zajmujących się siecią nie traktuje takich form religijności na serio i nie uwzględnia w swoich rozważaniach, bowiem są to parodie religii mające charakter satyryczny. Uwaga ta jest sporym nadużyciem, gdyż istnieje już bardzo dużo poważnych naukowych opracowań dotyczących religii cyberprzestrzeni (zob. Siuda 2010). Autorka z nich zresztą korzysta, czasami niezbyt trafnie odczytując intencje różnych badaczy.

Warto zaznaczyć, że wspomnianych powyżej interesujących rozważań znaleźć można w omawianym tytule bardzo dużo, tutaj wspomniana została zaledwie garstka. Każdego czytelnika dzieło zachwyci wieloma inspirującymi tezami związanymi z kulturami Internetu. Warto zatem udać się razem z Magdaleną Kamińską na polowanie na naukowo-kulturowo-internetowe memy, od których w książce aż się roi.

Bibliografia

Blackmore Susan (2000) *The Meme Machine*. Oxford, New York: Oxford University Press.

Siuda Piotr (2010) *Religia a Internet. O przenoszeniu religijnych granic do cyberprzestrzeni*. Warszawa: Wydawnictwa Akademickie i Profesjonalne.

Tarkowski Alek (2006) *Internet jako technologia i wyobrażenie. Co robimy z technologią, co technologia robi z nami?* [w:] Dominik Batorski, Mirosława Marody, Andrzej Nowak, red., *Społeczna przestrzeń internetu*. Warszawa: Academica, s. 20–34.

Cytowanie

Siuda Piotr (2014) Recenzja książki: „Magdalena Kamińska (2011) *Niechne memy: Dwanaście wykładów o kulturze Internetu*”. Poznań: Galeria Miejska Arsenal. „Przeгляд Socjologii Jakościowej”, t. 10, nr 2, s. 212–216 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przehladsocjologiijakosciowej.org>.

Kazimierz Kowalewicz
Uniwersytet Łódzki

DOI: <https://doi.org/10.18778/1733-8069.10.2.12>

Recenzja książki

Tomasz Ferenc (2012) *Artysta jako Obcy. Socjologiczne studium artystów polskich na emigracji*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, Wydawnictwo Biblioteki Państwowej Wyższej Szkoły Filmowej, Telewizyjnej i Teatralnej

Artysta i świat

Książka Tomasza Ferenc pod tytułem *Artysta jako Obcy. Socjologiczne studium polskich artystów na emigracji* jest dokonaniem imponującym, co uwidacznia się na płaszczyźnie logistycznej i w warstwie intelektualnej tekstu.

Ta pierwsza może w tej chwili wydawać się mniej ważna, ale właśnie dlatego należy poświęcić jej nieco uwagi. Powstało dzieło przygotowane poza ca-

łym dzisiejszym nurtem czy systemem grantów, stypendiów, programów i tym podobnych możliwości badawczych. Dzieło będące wynikiem autorskich umiejętności organizowania sobie badań, głęboko przeniknięte ideą prowadzenia *badania własnych* w najlepszym sensie tych słów, które – w powiązaniu z pasją badawczą i głęboką wiedzą – zaowocowały tekstem zasługującym na bardzo wysoką ocenę. Tutaj (chodzi o aspekt logistyczny) T. Ferenc może być szlachetnym przykładem, a jego książka stanowi też rodzaj zwierciadła, w którym odbić się mogą różne aspekty finansowania i prowadzenia badań czy szerzej – uprawiania nauki w naszym kraju.

O płaszczyźnie logistycznej nie trzeba byłoby wspominać, gdyby nie znakomita warstwa intelektualna czy poznawcza tekstu. Zresztą oczekiwania w zakresie osiągnięć intelektualnych w przypadku badań naukowych są czymś zwyczajnym, ale często spełnianym w różnym stopniu.

Otwierające książkę tytułowe słowo „artysta” mogłoby wskazywać, że autor powrócił do początków

swoich „poważnych” naukowych zainteresowań – wszak jego opublikowana rozprawa doktorska dotyczyła artystów. I choć to wszystko prawda, to takie skojarzenie jest do pewnego stopnia mylące. Jesteśmy bowiem w radykalnie innym miejscu analizy – to już nie artysta i jego sztuka są przedmiotem opisu, ale artysta, który jako człowiek znalazł się w sytuacji wyjątkowej – zdecydował się na emigrację. Ferenc mówi wtedy zazwyczaj, że chodzi o społeczną sytuację artysty. Ze względów czysto formalnych dodajmy, że recenzowana publikacja liczy trzysta osiemdziesiąt pięć stron plus dwie strony informacji od redakcji. Całość podzielona jest na 6 zasadniczych rozdziałów o różnej długości, poprzedzonych wstępem i domkniętym zakończeniem oraz bibliografią.

W krótkim, siedmiostronicowym wstępie autor ujawnia powody podjęcia tego właśnie tematu, stara się ten wybór uzasadnić i przedstawia w ogólnym zarysie treść poszczególnych rozdziałów. Przy okazji dokonuje pierwszych teoretycznych rozstrzygnięć, definiuje niektóre pojęcia, umiejętnie kreśli ścieżkę, którą nas poprowadzi. Ma też świadomość, że musi być otwarty na polifoniczność głosów, które na niej spotkamy i, uprzedzając nieco, zaznaczę, że robi to wyjątkowo udanie. Język Ferenc jest żywy, celny i precyzyjny. Sugestie, które płyną z dekonstrukcjonizmu, zostały przez autora *Artysty jako Obcego* odpowiednio przyswojone i przepracowane.

Rozdział pierwszy – zgodnie z określeniem zawartym w tytule – ma charakter przeglądowy i skoncentrowany jest na problematyce migracji, dyskusjach na ten temat, dalej – sprawach teorii i kon-

cepcjach badawczych pojawiających się w obrębie analiz związanych z wyróżnioną powyżej kwestią. Jest to syntetycznie opracowany rozdział, bardzo gęsty, choć zarazem przejrzysty, który wymaga od czytelnika skupienia uwagi. Ferenc maksymalnie wykorzystuje każdą stronę, by przedstawić to, co jest niezbędne w dalszej analizie. Nic jednak nie dzieje się poprzez proste odrzucenie czy zaniegowanie, zawsze chodzi o klarowność i jasność w sformułowaniu własnego stanowiska. Podział tego rozdziału i kolejnych (z wyjątkiem krótkiego fragmentu metodologicznego) na podrozdziały ułatwia lekturę.

Autor jest świadom problemów, które pojawiają się w badaniach nad procesami migracyjnymi. Jednym z nich jest napięcie pomiędzy tym, co mówią statystyki i co przynoszą studia indywidualnych przypadków. Kończąc ten rozdział, pisze: „[w] każdej indywidualnej biografii realizuje się zarówno to, co jednostkowe i niepowtarzalne, jak i to, co staje się udziałem grup doświadczających w zbliżonym czasie i okolicznościach zbliżonych sekwencji biograficznych” (s. 41). To właśnie wywiady biograficzne zebrane przez badacza w czterech metropoliach zachodniego świata (Paryż, Berlin, Londyn, Nowy Jork) stały się podstawą poczynionych w książce analiz i pozwoliły na powstanie studium artystów polskich na emigracji. Ta wspomniana powyżej umiejętność syntezy i dokonywania wyborów jeszcze wyraźniej manifestuje się w rozdziale metodologicznym – drugim rozdziale pracy. Przy ogólnym spojrzeniu może on wydać się zbyt krótki. I nawet jeśli taki jest, to w moim odczuciu zawiera wszystko, co niezbędne. Autor znajduje nawet miejsce, by poczynić uwagi

Kazimierz Kowalewicz, profesor Uniwersytetu Łódzkiego. Pracuje w Katedrze Socjologii Kultury Instytutu Socjologii Uniwersytetu Łódzkiego. Interesuje się teorią kultury.

Adres kontaktowy:

Katedra Socjologii Kultury, Instytut Socjologii
Wydział Ekonomiczno-Socjologiczny, Uniwersytet Łódzki
ul. Rewolucji 1905 r. nr 41, 90-214 Łódź
e-mail: kowalew@uni.lodz.pl

krytyczne wobec pewnych rozwiązań i jak zawsze przedstawić własne stanowisko. Materiału do analizy dostarczyły Ferencowi badania biograficzne – łącznie 61 wywiadów zebranych w wymienionych powyżej miastach. Ta liczba także świadczy o tym, że wybrana książka jest, jak już zaznaczono, dobrą robotą logistyczną. I tak, jak często to robi, również teraz Ferenc ostrzega nas przed niebezpieczeństwami metody biograficznej, stąd swoiste wsparcie w triangulacji. W centrum prowadzonych analiz pozostały podstawowe struktury procesowe biografii wypracowane przez F. Schützego. Autor, badając narrację, jak sam pisze, skupił się na tym, „co” jest opowiadane. Poprzez wielość biorących udział w badaniu twórców, obraz emigracyjnego życia ludzi związanych jeszcze w Polsce ze sztuką nabrał wielu wymiarów, znakomicie analitycznie wyeksponowanych w każdym z rozdziałów.

Kolejne cztery rozdziały są opisem biograficznych losów artystów, których Ferenc odnalazł – czy lepiej, do których dotarł – w czterech wielkich światowych stolicach, miastach na różne sposoby sprzęgniętych także z rozwojem sztuki nowoczesnej.

Rozdział trzeci przedstawia procesy asymilacyjne i strategie artystyczne polskich twórców w Paryżu. O takiej kolejności „opowiadania” o losach artystów zdecydowało zapewne miejsce Paryża w dziejach sztuki i historii sztuki polskiej oraz polskiej emigracji. Rozdział jest dość obszerny, ale jak zwykle dobrze ustrukturuwany. Choć we wprowadzeniu autor przypomina historię emigracji Polaków do Francji, to same narracje dotyczą końca lat sześćdziesiątych, dalej kroczymy poprzez doświadczenia roku 1981 i lat następnych, aż po cza-

sy współczesne, kiedy każdy z emigrantów może do kraju swobodnie wjechać i wyjechać. Nie każdy się jednak na to decyduje.

Pisząc ten rozdział poprzez rzeczywiste jednostkowe losy, autor – zgodnie z zapowiedzią – chce odnaleźć to, co jest udziałem grup doświadczających podobnych sekwencji biograficznych. I to się w pełni powiodło. Przekonuje nas o tym podsumowanie, w którym korzystając z propozycji socjologii sztuki i socjologii skupionej na kwestiach tożsamości, Ferenc stara się określić pozycję/miejsce polskiego artysty-emigranta we Francji. Wydaje się, że dla opisu tej sytuacji najlepsze jest pojęcie Obcego, który na „zawsze pozostaje pomiędzy” (s. 129), niby włącza się w nową kulturę, ale nie robi tego do końca. Ale to „pomiędzy” jest częściowo wymuszone przez same zachowania przedstawicieli kultury przyjmującej, a po części jest chyba śladem – tak się wydaje – sarmackiej tradycji, która ciąży na nas, naszych rodakach także z dala od kraju. Może również moment, w którym emigracja miała miejsce, był nie bez znaczenia.

Do tego tematu czy wątku „pomiędzy” autor powraca w znacznie bardziej rozbudowanej postaci – kiedy w rozdziale czwartym pisze o emigracji do Wielkiej Brytanii. Nie mógł, także w tym przypadku, ze względu na historyczne zaszłości, pominąć całej historii emigrowania do tego kraju i roli tej emigracji w samej historii Polski, a także nie zauważyć różnic pomiędzy kolejnymi pokoleniami emigrujących. By dobrze opisać zjawisko emigracji do Wielkiej Brytanii, Ferenc wyróżnił właściwie cztery fale migracji. Najciekawsza jest ostatnia z emigracyjnych fal. To chyba do tych, którzy

ostatnio wyjechali, odnosi się słowo „pomiędzy”, ale opisywane teraz przy pomocy języka, jakiego dostarczyła koncepcja V. Turnera. Badacz traktuje doświadczenia tej ostatniej fali jako doświadczenie liminalne. Ale, co autor zaznacza, takie doświadczenia nie są obce również emigrującym polskim artystom z Paryża czy z Nowego Jorku. Owa liminalność odciska się czasem także na tożsamości danego artysty – jest ona wtedy tożsamością liminalną. W narracjach zauważyć można różne inne tożsamości. Część z nich realizuje ponowoczesny wariant tożsamości określony przez Z. Baumanianem turystą.

Inni emigranci z kolei nie uniknęli wariantu włóczęgi, ale, jak sugeruje Ferenc, warto jeszcze pamiętać o wędrowcy, który nie jest ani turystą, ani włóczęgą. Nie bez powodów wyróżnił on poszczególne fale emigracji, gdyż to one w pewien sposób odciskają się na wariantach manifestowanych tożsamości. Nie sposób zresztą krótko omówić całego bogactwa problemów, które poruszone są w tym rozdziale. Dałem zaledwie krótki szkic ważniejszych zagadnień.

Najobszerniejszy jest rozdział piąty – poświęcony polskim artystom w Nowym Jorku, liczy ponad sto stron. Jak zwykle jest on dobrze ustrukturuwany i zarazem bardzo gęsty. Przynosi rozważania skupione wokół czterech zasadniczych procesów biograficznych, wyróżnionych w klasycznej już literaturze przedmiotu w następstwie analiz wywiadów narracyjnych. Po trosze ze stereotypową formułą dotyczącą Stanów Zjednoczonych w pamięci, jesteśmy w obszarze wyznaczonym przez dwa słowa: sława i rozpacz. Kiedy okazało się, że to Nowy Jork

jest stolicą sztuki nowoczesnej, miasto stało się celem wędrowki także polskich artystów. Ich zmagania z nowym miejscem możemy śledzić w wielu wymiarach, poznajemy drogi prowadzące do sukcesu, dowiadujemy się o tożsamości wędrowców czy też poznajemy losy tych, którzy znaleźli się na trajektoryjnej ścieżce i przestali decydować o sobie, swoim losie i własnej tożsamości. Ale mimo całej odmienności codziennych doznań i kłopotów czy też sukcesów, może zaskoczyć nas zdanie z podsumowania piątego rozdziału „[b]iograficzne opowieści nowojorskich imigrantów posiadają wiele elementów wspólnych z narracjami twórców, którzy osiedlili się w Europie” (s. 306). Nie bez znaczenia dla konstruowania tożsamości artysty – emigranta w Nowym Jorku – jest poczucie obcości, które popycha do poszukiwań tożsamości. Ferenc proponuje, by dla części tych zmagających zastosować termin wielotożsamości, objawiającej się poprzez „współwystępowanie wielu punktów identyfikacyjnych ważnych dla jednostki, które dynamicznie wpływają na siebie i poddawane są nieustannej refleksyjności” (s. 310). Może być ona dostępna tym, którzy po prostu zapragną pozostać obcy, dodaje autor.

Najkrótszym z rozdziałów o emigrujących artystach jest rozdział szósty – opisujący emigracyjne losy Polaków w Berlinie. Tak jak poprzednio, perspektywa historyczna przenika się w nim z obserwacjami i aktualnymi doświadczeniami. Oprócz narracji artystów, którzy się na emigrację zdecydowali, poznajemy opowieści tych, którzy musieli wyjechać wraz z rodzicami i na emigracji dorastali. Podobnie jak w przypadku artystów z Londynu, oddzielne miejsce poświęcono tym, którzy

wyemigrowali po 1989 roku i później. Te narracje kontrastują z tymi, które słyszymy od osób opuszczających wcześniej kraj z powodów politycznych. Tutaj – w swoim czasie – szczególnie miejsce zajmował Berlin Zachodni – jako część wolnego zachodniego świata był ważnym ośrodkiem niezależnej sztuki i miejscem dążeń także twórców z Polski. Ten rozdział jest rzeczywiście niedługi, ale być może w samym mieście ukryty jest potencjał ważny dla poszukujących zarówno miejsc dla swojej twórczości, jak i innych sposobów życia. W jakimś zakresie można zauważyć, że to do Berlina najczęściej prowadzą ostatnio drogi emigrujących polskich artystów. A przy tym, jak pisze Ferenc, jest to miejsce, które zmusza do działania. Zatem choć ten rozdział miał we wstępnej części wydźwięk dramatyczny, to jego zamknięcie jest optymistyczne.

Ale to wcale nie oznacza, że to samo możemy pomyśleć o losach tych, którym poświęcona jest książka i których poza Berlinem autor spotykał w Paryżu, Londynie czy Nowym Jorku.

Jakie są zatem podstawowe zalety i badawcze ustalenia przedstawione w recenzowanej publikacji? W każdym miejscu, na dowolnej stronie czytelnik może zauważyć, że jest to prawdziwie wielowątkowe i gruntowne socjologiczne studium dotyczące polskich artystów, którzy zdecydowali się na emigrację. Najważniejsze pozycje literaturowe autorów polskich i obcych stanowią dla Ferenc'a punkt odniesienia, prowadzi z nimi dialog, czasem przeraża się on w spór i wreszcie propozycję własnego stanowiska. Tak jest w każdym rozdziale, najwyraźniej zaś w tych, które poświęcone są opisowi życia artystów-emigrantów w czterech wybranych

wielkich miastach zachodniego świata. Dzieje się tak niezależnie od długości rozdziału, użytych kategorii czy wreszcie próby podsumowania – przedstawienia „obrazu” danego miejsca.

To wszystko nie byłoby możliwe, gdyby nie wyjątkowe badania terenowe zrealizowane przez Ferenc'a. Bardzo wysoko należy ocenić nie tylko ich zakres, ale i umiejętność prowadzenia analizy zebranych wywiadów i wreszcie autorską narrację. Mówiąc językiem dzisiejszego „komputerowego” świata, napotykamy w tekście dziesiątki „linków”, nad którymi autor bezwzględnie panuje. A przy uważnej lekturze, panuje nad nimi także czytelnik.

Zawsze w końcu, mimo zmiennych nazw podrzdziałów, czasem ich różnej długości, nieraz wybranych kategorii analizy, możemy porównać losy ludzi, których spotkać możemy w Berlinie, Paryżu, Londynie czy Nowym Jorku, ale obecnie coraz częściej też na polskiej ulicy. Kiedy szukam dobrej miary dla dokonania oceny tej książki, choć to może wydać się dziwne i zaskakujące, jedynym i właściwym wydaje mi się dzieło W. I. Thomasa i F. Znanieckiego pod tytułem *Chłop polski w Europie i Ameryce*. Mimo wszystko, zaryzykuję to porównanie z wielkim klasycznym dziełem i dodam, że znam oczywiście proporcje, więcej – powinienem i staram się je zachować. Ale tak, jak praca Thomasa i Znanieckiego ma obecnie wszelkie cechy dzieła klasycznego, tak nabywać ich z pewnością będzie książka T. Ferenc'a. Co się ostatecznie stanie, nie mnie przesądzać, choć niemiło jest się mylić.

W zakończeniu tej recenzji, by uniknąć wszelkich wątpliwości, stwierdzam, że omawiana publikacja

wnosi wkład w rozwój dyscypliny naukowej, jaką jest socjologia. Dotyczy on wszechstronnego i pogłębionego socjologicznego ujęcia procesów emigracji artystów polskich do wybranych miast Europy i Ameryki. Jednocześnie, poprzez cytowanie materiałów biograficznych i autorskie komentarze oraz analizy, czytelnik może zupełnie inaczej spoj-

rzeć na funkcjonowanie społeczeństw, do których dotarli emigranci. To zatem swoisty przyczynek do rozbudowanego socjologicznego ich opisu. Z pewnością nie powinni go również pomijać socjologowie z tych państw, do których prowadziły drogi polskich artystów, nawet jeśli nie interesują się sztuką.

Cytowanie

Kowalewicz Kazimierz (2014) *Recenzja książki: Tomasz Ferenc (2012) „Artysta jako Obcy. Socjologiczne studium artystów polskich na emigracji”*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, Wydawnictwo Biblioteki Państwowej Wyższej Szkoły Filmowej, Telewizyjnej i Teatralnej. „Przegląd Socjologii Jakościowej”, t. 10, nr 2, s. 218–223 [dostęp dzień, miesiąc, rok]. Dostępny w Internecie: <www.przegladsocjologiijakosciowej.org>.



PTS jest naukową i zawodową organizacją socjologów polskich. Działa od 1957 roku. Liczy około 1330 członków. Posiada oddziały terenowe w:

Białymstoku,	Łodzi,	Toruniu,
Gdańsku,	Opolu,	Warszawie,
Katowicach,	Poznaniu,	Wrocławiu,
Krakowie,	Rzeszowie,	Zielonej Górze
Lublinie	Szczecinie	

oraz sekcje tematyczne:

Antropologii Społecznej,	Socjologii Miasta, Socjologii Wsi,
Historii Socjologii,	Socjologii Młodzieży i Edukacji,
Kół Naukowych,	Socjologii Nauki, Socjologii Pracy,
Metodologii Badań Społecznych,	Socjologii Prawa,
Pracy Socjalnej,	Socjologicznych Problemów Bezpieczeństwa Narodowego,
Socjologii Dewiacji i Kontroli Społecznej,	Socjologii Religii,
Socjologii Jakościowej i Symbolicznego Interakcjonizmu,	Socjotechniki
Socjologii Medycyny,	

Co trzy lata Towarzystwo organizuje ogólnopolskie zjazdy socjologiczne, które podsumowują wiedzę o społeczeństwie i wskazują nowe kierunki badań.

Pod auspicjami PTS ukazują się: anglojęzyczny kwartalnik „Polish Sociological Review” i biuletyn, wydawany dwa razy do roku, adresowany do członków PTS „Informacja Bieżąca PTS” oraz „Bibliograficzna Informacja Bieżąca”, która rejestruje poważną część prac socjologicznych wydawanych w Polsce.

Od 1974 roku PTS wyłania w drodze konkursu najlepszą książkę napisaną przez autora młodszego pokolenia i przyznaje prestiżową Nagrodę im. Stanisława Ossowskiego. Od lat 90. przyznawana jest także nagroda im. Floriana Znanieckiego za najlepsze prace magisterskie z socjologii. Od 2010 roku PTS przyznaje również nagrodę za najlepsze tłumaczenie książki socjologicznej.

Od chwili swego powstania PTS jest kolektywnym członkiem ISA i współpracuje z Europejskim Stowarzyszeniem Socjologicznym (ESA).

W 2005 roku PTS uzyskał status organizacji pożytku publicznego.

Polskie Towarzystwo Socjologiczne
ul. Nowy Świat 72, p. 216
00-330 Warszawa
tel./fax (0-22) 826-77-37
e-mail: pts@ifispan.waw.pl
www.pts.org.pl

PSJ

Dostępny online
www.przegladsocjologiijakosciowej.org

„PSJ” stworzyliśmy, aby umożliwić swobodny przepływ informacji w społecznym świecie socjologii jakościowej. Adresujemy go do wszystkich socjologów, dla których paradygmat interpretacyjny i badania jakościowe stanowią podstawową perspektywę studiowania rzeczywistości społecznej.

Analiza danych jakościowych wspomagana komputerowo

Tom X ~ Numer 2

31 maja 2014



*Publikacja dofinansowana ze środków
Polskiego Towarzystwa Socjologicznego
Seria Wydawnicza PTS, Edycja IV
– „Co po kryzysie?”*

REDAKTOR NUMERU TEMATYCZNEGO: Jakub Niedbalski

REDAKTOR NACZELNY: Krzysztof T. Konecki

REDAKCJA: Waldemar Dymarczyk, Marek Gorzko
Anna Kacperczyk, Sławomir Magala
Łukasz T. Marciniak, Jakub Niedbalski
Izabela Ślęzak

KOREKTA I EDYCJA: Magdalena Chudzik,
Magdalena Wojciechowska

PROJEKT OKŁADKI: Anna Kacperczyk

ISSN: 1733-8069