




Anna Majewska-Tworek
Uniwersytet Wrocławski
 <https://orcid.org/0000-0003-3997-5991>
anna.majewska-tworek@uwr.edu.pl

Forum Lingwistyczne
nr 7, 2020
ISSN 2450-2758 (wersja elektroniczna)



DOI: <http://doi.org/10.31261/FL.2020.07.06>

Monika Zaśko-Zielińska
Uniwersytet Wrocławski
 <https://orcid.org/0000-0001-6333-660X>
monika.zasko-zielinska@uwr.edu.pl

Piotr Pęzik
Uniwersytet Łódzki
 <https://orcid.org/0000-0003-0019-5840>
piotr.pezik@uni.lodz.pl

„Polszczyzna mówiona miast” – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej

“Spoken Polish of cities” – a Continuation of Research from the 1980s
with the Use of Digital Linguistics Toolss

Abstract: The subject of the research are old texts recorded as part of the “Spoken Polish of cities” project, which was carried out in Katowice, Kraków, Łódź and Wrocław in the last decades of the 20th century. The main aim of the article is to present a new study of Wrocław oral data from that period. The tools of digital humanities and corpus linguistics were used in the new study. The result of the team’s work is the *Speech of Wrocław* collection in Spokes resources.

Key words: spoken language, sociolinguistics, corpus, Spokes

Abstrakt: Przedmiotem badań są stare teksty nagrane w ramach projektu „Polszczyzna mówiona miast”, który był realizowany w Katowicach, Krakowie, Łodzi i we Wrocławiu w ostatnich dekadach XX wieku. Głównym celem artykułu jest przedstawienie nowego opracowania wrocławskich danych mówionych pochodzących z tamtego okresu. Do nowego opracowania zastosowano narzędzia humanistyki cyfrowej oraz językoznawstwa korpusowego. Rezultatem pracy zespołu jest zbiór *Mowa Wrocławia* w zasobach Spokes.

Słowa kluczowe: mówiona odmiana języka, socjolingwistyka, korpus, Spokes

1. Wstęp

Głównym powodem powstania niniejszego artykułu jest odnalezienie w Instytucie Filologii Polskiej Uniwersytetu Wrocławskiego kaset z nagraniami, które były gromadzone na potrzeby dawnego projektu, bardzo znanego w całym środowisku polonistycznym, poświęconego badaniom polszczyzny mówionej miast w latach 70., 80. i 90. XX wieku. Informacja o znalezionym materiale spotkała się z zainteresowaniem pracowników polskiego zespołu CLARIN, dzięki czemu materiał został ponownie opracowany.

Nikt jednak nie szukałby kaset ani na nowo ich nie transkrybował, gdyby nie swoisty renesans zainteresowań polszczyzną mówioną w różnych nurtach współczesnego językoznawstwa teoretycznego i stosowanego, choć znawcy przedmiotu zwracają uwagę na brak odpowiednio przygotowanych, wystarczająco obszernych materiałów z niezbędnymi metadanymi. Istnieje zatem wielka potrzeba wzbogacenia istniejących korpusów wypowiedzi mówionych oraz tworzenia nowych, by można było z wykorzystaniem nowoczesnych narzędzi badawczych rzeczywiście zintensyfikować dociekanie istoty mówionej odmiany języka w różnych jej przejawach. Nie sposób tego robić bez odwołania się do początkowych ustaleń badaczy z Katowic, Krakowa, Łodzi i Wrocławia, ich doświadczeń oraz osiągnięć.

2. Polszczyzna mówiona miast w XX wieku – początki projektów

Badania nad polszczyzną mówioną mieszkańców Katowic prowadzone przez badaczy z Uniwersytetu Śląskiego w Katowicach pod kierunkiem Władysława Lubasia rozpoczęły się już w pierwszej połowie lat 70. Później objęły cały obszar ówczesnego województwa katowickiego. W czerwcu 1974 roku w Sosnowcu z inicjatywy pracowników Instytutu Filologii Polskiej Uniwersytetu Śląskiego w Katowicach odbyła się ogólnopolska konferencja naukowa poświęcona potocznej polszczyźnie mówionej miast. W obradach uczestniczyli przedstawiciele większości polonistyk polskich. Rezultatem tego spotkania było wydanie w 1976 roku publikacji zatytułowanej *Miejska polszczyzna mówiona. Metodologia badań*. We wstępie do tomu Lubaś podkreślił konieczność zespołowej pracy nad badaniem polszczyzny mówionej oraz znaczenie wspólnej metodologii wynikającej z polskich i zagranicznych doświadczeń. Zaznaczył także, że opublikowany zbiór tekstów ma szansę zainicjować dalszy rozwój zainteresowań naukowych polszczyzną mówioną (LUBAŚ, 1976: 6–8), której wcześniejsze badania zostały w tomie zrelacjonowane w artykułach Bronisława Wieczorkiewicza, Mieczysława Karasia i Alfreda Zaręby (LUBAŚ, 1976: 9–40).

Pod koniec 1976 roku z inicjatywy Karasia został zainaugurowany projekt krakowski, który wyróżniał się uwzględnieniem wielu pozajęzykowych zmiennych już na etapie jego planowania (DUNA, 1981: 7). Również w 1976 roku rozpoczęły się prowadzone pod kierunkiem Marii Kamińskiej badania polszczyzny mówionej mieszkańców Łodzi. Tutaj postanowiono, że pierwszym zadaniem będzie gromadzenie materiału, a dopiero później na tej podstawie zostaną sformułowane pewne ustalenia teoretyczne (KAMIŃSKA, 1979: 87). Dopiero w latach 80. do badania polszczyzny mówionej przystąpili wrocławianie, chociaż wcześniejsze prace Antoniego Furdala nad dialektologią miejską (FURDAL, 1966) oraz kwestionariusz do badania dialektów kulturalnych i miejskich (FURDAL, 1973) często stawały się punktem wyjścia do rozważań metodologicznych prowadzonych w innych ośrodkach naukowych.

3. Szczegółowe założenia wstępne. Obszar badań i dobór populacji

Mimo że w kontekście języka miasta problem obszaru badań mógłby się wydawać oczywisty, to zarówno na etapie konstruowania planów zbierania materiału, jak i w refleksjach

późniejszych okazało się, że jest to zagadnienie wymagające dyskusji i doprecyzowania. W opublikowanej w 2000 roku pracy poświęconej typologii współczesnej polszczyzny Aleksander Wilkoń wprowadził pojęcie miejskich języków mieszanych, które na podstawie istniejących już wtedy badań nad polszczyzną mówioną miast podzielił na kilka typów (WILKOŃ, 2000: 28–30). Przykładem jednego z tych typów były miejskie dialekty występujące w miastach górnośląskich (np. Katowicach), w których język ogólny i gwara istnieją obok siebie, tworząc charakterystyczny dwujęzyczny obszar językowy. Natomiast Kraków reprezentuje typ organizmu miejskiego, w którym dominuje język ogólny, ale na obrzeżach miasta (dawne wsie podkrakowskie) język zawiera wiele cech regionalnych. Jeszcze inny typ mogłaby reprezentować Łódź, która, podobnie jak opisany przez Wilkońa Szczecin, przedstawia obszar pewnej integracji językowej.

Hipoteza o jednorodności językowej przestrzeni miasta czasem skutkowałą wnioskiem, że podczas zbierania materiału stosowanie podziału terytorialnego w obrębie organizmu miejskiego jest zbędne. Takie podejście charakterystyczne było dla badań nad językiem miejskim Katowic. Badaniami objęto ówczesne terytorium Katowic bez podziałów przestrzennych (np. na centrum i przedmieścia), gdyż założono, że cały ten obszar łączą więzi społeczne. Jednak oprócz materiałów pochodzących z Katowic w ramach projektu zebrano także teksty z kilkunastu miast śląskich i zagłębiowskich. Wśród nich znalazły się: Gliwice, Bytom, Siemianowice Śląskie, Mikołów, Mysłowice, Sosnowiec, Czeladź, Będzin, Dąbrowa Górnicza, Siewierz, Chorzów, Ruda Śląska, Zawiercie, Rybnik, Świętochłowice, Tarnowskie Góry, Tychy, Wodzisław Śląski, Zabrze (LUBAŚ, 1978; 1980).

Na terytorium Krakowa badania były prowadzone w granicach sześciu dzielnic: Grzegórzki, Kleparz, Nowa Huta, Podgórze, Stare Miasto, Zwierzyniec. Uwzględniono zatem podział miasta obowiązujący w latach 1952–1973. Wiadomo było bowiem, że każda z dzielnic ma swoją odrębną strukturę socjologiczną, która może mieć wpływ na cechy językowe jej mieszkańców.

Z kolei w ośrodku łódzkim badania prowadzono w granicach Wielkiej Łodzi, ale z dbałością o równowagę danych podzielono miasto na pięć części (północną, wschodnią, centralną, zachodnią, południową) z dokładnym wyznaczeniem dzielnic, które wchodziły w ich skład (KAMIŃSKA, 1979: 89). Jeszcze przed przystąpieniem do badań przypuszczano, że raczej nie istnieje żadna odrębna łódzka gwara miejska. Zakładano również, że z powodu centralnego położenia miasta na mapie Polski oraz sąsiedztwa gwar łęczycko-sieradzkich, które nie wyróżniają się wyjątkowymi odrębnościami, w mieście „winna się wykształcić nadgwarowa odmianka języka mówionego – swoisty interdialekt” (KAMIŃSKA, 1979: 87).

We Wrocławiu postanowiono prowadzić badania w granicach administracyjnych miasta, zwracając jednocześnie uwagę na dzielnice o charakterze wiejskim i półwiejskim (słabo zintegrowane z centrum organizmu miejskiego) oraz intensywne uczestniczenie w życiu miasta osób dojeżdżających z niedalekich miasteczek, na przykład z Oławy, Trzebnicy, Kątów Wrocławskich, Obornik Śląskich. Podkreślano, że miasta te nigdy nie były prowincją względem wrocławskiej aglomeracji, stąd nie tak wielkie różnice socjolingwistyczne występowały między ich mieszkańcami jak w przypadku aglomeracji krakowskiej czy łódzkiej. Ponadto Franciszek NIECKULA pisał: „Nie ma [...] żadnych powodów, by w jakimkolwiek stopniu liczyć się z podziałem aglomeracji na 5 dzielnic: Stare Miasto, Śródmieście, Krzyki, Fabryczną i Psie Pole. Podział ten nie odpowiada żadnym realiom historycznym, architek-

tonicznym, komunikacyjnym czy społecznym” (1990: 24)¹. Po wojnie Wrocław wyróżniał się przede wszystkim wielojęzyczną mozaiką. Ogromne zniszczenie miasta (Wrocław w wyniku działań wojennych był zburzony w 70%) powodowała „ciągłą migrację wewnątrz organizmu miejskiego [...]. Wszystko to znakomicie sprzyjało integracji i utrudniało wyłanianie się wyraźnych nowych zróżnicowań socjolingwistycznych” (NIECKULA, 1990: 24).

Drugim ważnym problemem w badaniach polszczyzny mówionej miast był dobór populacji i ustalenie reprezentatywnej próby badawczej. W odniesieniu do Katowic – ze względu na to, że tu za najważniejsze kryterium uznano terytorium – ustalono, iż badaną populację będą tworzyli zarówno rodowici katowiczanie, osoby zamieszkałe w Katowicach, jak i ci, którzy mieszkają poza miastem, ale pracują w Katowicach przynajmniej od roku.

W Krakowie za konieczne uznano odróżnienie osób urodzonych w badanym mieście (wśród tych osób skrupulatnie odnotowywano miejsce urodzenia rodziców, sprawdzając tym samym stopień zasiedlenia rodziny w mieście) od przybyszów (za ważną granicę czasową świadczącą o zrośnięciu się ze środowiskiem krakowskim uznano okres 10 lat), a także ludności przepływowej, tzn. mieszkających w mieście okresowo (studenci) albo też dojeżdżających do szkół czy zakładów pracy.

W Łodzi ze względu na nieustanny napływ nowej ludności postanowiono, że informatorami staną się osoby, które mieszkają w mieście przynajmniej 12–15 lat, gdyż dopiero tak długi pobyt pozwala na asymilację ze środowiskiem miejskim (KAMIŃSKA, 1979: 89).

Prawie całkowita wymiana ludności w powojennym Wrocławiu spowodowała, że w doborze populacji trzeba było uwzględnić pochodzenie przybyszów oraz czas zamieszkania w mieście. W badaniach wrocławskich planowano skupić się na trzech grupach wiekowych. Pierwszą miała stanowić najstarsza generacja, czyli osoby podeszłeścdziesięcioletnie (w 1945 r. miały 20 lat i więcej); drugą grupę – osoby urodzone i wychowane we Wrocławiu, czyli wrocławianie trzydziestoletni i czterdziestoletni; trzecią grupę – dzieci w wieku szkoły podstawowej (7–15 lat). Liczono więc na przebadanie trzech generacji, dzięki czemu udałoby się zaobserwować pewne zmiany językowe w mowie najmłodszych: „[...] w ich mowie mogą się ujawniać, choć nie muszą, cechy językowe najstarszej generacji (dziadkowie, babcie), właściwości drugiej grupy (rodzice) oraz zapowiedź tej polszczyzny, która za jakiś czas będzie dominowała we Wrocławiu jako ostateczny rezultat procesów integracyjnych” (NIECKULA, 1990: 31).

Kolejnym ważnym zagadnieniem w opisywanych projektach był dobór badanych. Tu znów poszczególne ośrodki nieco różniły się w szczegółowych ustaleniach. Spowodowane to było możliwościami organizacyjnymi czy wielkością zespołu, a także specyfiką miejsca. Cechę wspólną stanowiła rezygnacja z doboru losowego mówców jako mało efektywnego zarówno organizacyjnie, jak i merytorycznie.

W badaniu mieszkańców miast śląskich początkowo w doborze populacji uwzględniano przede wszystkim czynniki socjalne: zawód, wykształcenie, etniczność regionalną, wiek, płeć oraz bilingwizm. Później zwrócono uwagę, że ten zespół czynników ma różną istotność w poszczególnych grupach społecznych. Dlatego też rozważano zastąpienie kryterium socjalnego „typem kultury”, na który składałyby się: uwarunkowana historycznie etniczność,

¹ Chociaż powszechnie wiadomo, że były dzielnice szczególnie często i chętnie zamieszkiwane przez inteligencję, np. Biskupin.

ograniczany pozycją zawodową zakres kontaktów społecznych, właściwości osobowe powiązane z predyspozycjami psychicznymi, wiek, płeć oraz wykształcenie (LUBAŚ, 1976: 44).

W Krakowie przed podjęciem badań lingwistycznych wytypowano we współpracy z socjologami czynniki socjalne i biologiczne, które mogą mieć wpływ na różnice językowe w środowisku miejskim. Były to: wiek, płeć, miejsce urodzenia, pochodzenie społeczne, wykształcenie, zawód oraz charakter wykonywanej pracy (DUNAJ i in., 1979: 80–81; por. też DUNAJ, 1987: 36). Dzięki przystąpieniu do próbnych analiz fonetycznych, leksykalnych i składniowych (czyli do „wstępnej konfrontacji z materiałem” – DUNAJ, 1981: 9) sprawdzono, że opracowana siatka zróżnicowań socjalnych Krakowa ma charakter maksymalny. Wywnioskowano też, że do badań zagadnień szczegółowych wystarczy uwzględnić tylko niektóre czynniki.

W Łodzi do badań mogły być wytypowane osoby, które mieszkają w mieście dłużej niż 10 lat. Postanowiono nie brać pod uwagę repatriantów z za Buga ani obcokrajowców. Postanowiono też skupić się na polszczyźnie łódzian zatrudnionych w przemyśle włókienniczym: „[...] łódzkie tkaczki i prządki oraz robotnicy zawodów pokrewnych mogą być uznani za najbardziej reprezentatywnych przedstawicieli łódzkiej społeczności” (KAMIŃSKA, 1979: 88). Natomiast badania łódzkiej inteligencji odsunięto na dalszy plan. Takie ograniczenie wynikało z chęci urealnienia całego projektu w rozsądnym przedziale czasowym. Uznano bowiem za niemożliwe przebadanie całego miasta w kilkuletnim projekcie badawczym.

W badaniach wrocławskich Nieckula wymienił kilkanaście zmiennych, które miały być uwzględniane w opisie zgromadzonych nagrań. Wydaje się, że nie stanowiły one listy zamkniętej. Przypuszczano, że nie wszystkie muszą być istotne podczas badania nagranych wypowiedzi. Dotyczyły one charakterystyki socjolingwistycznej mówców: wiek partnerów, płeć, wykształcenie, zakład pracy i zajmowane stanowisko, pochodzenie społeczne, zamożność, rodzaj zainteresowań, aktywność społeczna (działalność w stowarzyszeniach); charakterystyki nagranej wypowiedzi: tematyka wypowiedzi, forma tekstu (rozmowa, opowiadanie itd.), rodzaj kontaktu językowego (relacje społeczne między partnerami, oficjalność lub nieoficjalność), sytuacja aktu komunikacji (miejsce, czas itp.); charakterystyki psycholingwistycznej nagrywanych: łatwość werbalizacji myśli, tempo mówienia, dykcja; a także „czynników typowo wrocławskich”: czas pobytu w mieście, pochodzenie regionalne (i ewentualnie społeczne) rodziny, utrzymywanie (lub niepodtrzymywanie) więzi z miejscem pochodzenia rodziny (NIECKULA, 1990: 21).

4. Cele i metody badawcze

Celem projektu w Katowicach był opis języka miasta uwzględniający wszystkie jego składniki oraz wariantywność uwarunkowaną społecznie i kulturowo. Uznano, że koncentrowanie się na dialektyzmach lub języku ogólnym byłoby eksponowaniem jednego elementu całego mieszanego systemu, jakim jest język miejski. Spośród trzech typów kontaktów językowych charakterystycznych dla mieszkańców miast postanowiono skoncentrować się głównie na kontaktach lokalnych oraz indywidualnych zachodzących w małych grupach w sytuacjach rodzinnych i przyjacielskich. W tego typu relacjach ujawnia się bowiem najbardziej wariantywność języka.

W Krakowie, formułując metody prowadzenia badań miejskiej polszczyzny mówionej, często odwoływano się do doświadczeń badań dialektologicznych. Jednak celem dialektologów – jak słusznie podkreślano – było zazwyczaj dotarcie do najstarszej formy mówionej języka na danym terenie. Z tego powodu badano przedstawicieli (nawet pojedynczych) najstarszego pokolenia. Z kolei zasadniczy cel badań tzw. dialektologii miejskiej miał być inny: „[...] idzie w nich bowiem o pokazanie zróżnicowania językowego jakiegoś środowiska, uwarunkowanego rozmaitymi czynnikami społecznymi i sytuacyjnymi. Z tego względu wybór reprezentatywnych informatorów staje się przedsięwzięciem znacznie bardziej skomplikowanym” (DUNAJ, 1987: 30). Dużym wyzwaniem badawczym była więc nie tylko decyzja o doborze próby badanych, ale także określenie techniki zbierania materiału oraz wyznaczenie sytuacji, w których to się będzie odbywało (DUNAJ, 1987: 30). W badaniach nad polszczyzną mówioną Krakowa zastosowano metodę celowego doboru reprezentacji². Polega ona na celowym doborze próby „ze względu na określone cechy socjalne i liczbę osób potrzebnych do danej grupy” (DUNAJ, 1987: 33). W przeciwieństwie do doboru losowego umożliwia uzyskiwanie zrównoważonych danych (proporcjonalnych liczbowo) od poszczególnych, wcześniej starannie wyznaczonych grup socjalnie jednorodnych. W tego typu badaniach kolejnym dużym wyzwaniem jest określenie liczebności wyznaczonej grupy. Dunaj, powołując się na badania Haliny Kurek (1981), podał liczbę grupy minimalnej: „[...] odpowiednie obliczenia wykorzystujące rachunek prawdopodobieństwa dowodzą, że w badaniach dużej populacji grupy jednorodne socjalnie winny liczyć powyżej 10 osób” (DUNAJ, 1987: 34).

W Łodzi głównym celem badań był opis współczesnej polszczyzny łódzkiej, w tym przeprowadzenie analiz porównawczych między pokoleniami w obrębie wybranych rodzin robotniczych: „[...] opis polszczyzny łódzkiej winien być dokonany zasadniczo w płaszczyźnie synchronicznej, nie rezygnujemy jednak także w miarę możliwości z badań sięgających w przeszłość” (KAMIŃSKA, 1979: 88). Zauważono bowiem, że w mowie najstarszego pokolenia są jeszcze cechy polszczyzny dziewiętnastowiecznej, które koniecznie chciano odnotować. Pierwszym celem było przygotowanie zbioru tekstów mówionych mieszkańców Łodzi. KAMIŃSKA (1979: 89) nazwała tak sformułowane cele „minimalistycznymi”. Miały być one podstawą kolejnych badań. Zastosowano celowy dobór badanych: „[...] podobnie jak to się zwykło czynić przy badaniach dialektologicznych stosujemy staranny dobór informatorów, nie są oni typowani mechanicznie” (KAMIŃSKA, 1979: 90). Poszukiwano mówców chętnych do współpracy, o znanym miejscu urodzenia, miejscu zamieszkania i znanym zawodzie. Do zbioru danych nie włączano form zasłyszanych. Dość ściśle określono liczbę badanych. Z każdej części miasta (z 5 wcześniej wyznaczonych) postanowiono uzyskać co najmniej 100 tekstów. „W sumie z całego miasta byłoby to 500 dłuższych wypowiedzi, które nie wyznaczają liczby przebadanych informatorów. [...] Liczba informatorów winna sięgnąć 800 osób, co dawałoby w rezultacie 1 na 1000 mieszkańców miasta” (KAMIŃSKA, 1979: 89).

We Wrocławiu już na wstępie badań odrzucono hipotezę o istnieniu dialektu wrocławskiego (wrocławskiej odmiany języka polskiego). Zasadniczym celem było opisanie stanu współczesnej polszczyzny w tym mieście oraz poznanie zależności zjawisk językowych od zjawisk socjalnych i kulturowych (najlepiej także na tle innych dużych miast). Podobnie

² Opis losowego i celowego doboru próby badanych przedstawił DUNAJ (1987: 32).

jak w innych ośrodkach – i tu szczególnym zainteresowaniem badawczym cieszyła się nieoficjalna polszczyzna mówiona. Natomiast celem wyróżniającym wrocławskie badania miało być śledzenie procesów integracji językowej: „[...] w wyniku tak strasznego zdarzenia historycznego, jakim była II wojna światowa i związane z nią migracje ludności na skalę niespotykaną w dziejach, doszło do wytworzenia się zupełnie nowych społeczności wielkomijskich z elementów przynależnych do wszelkich możliwych regionów i grup społecznych. [...] Opis procesów integracyjnych i zebranie dla nich bogatej dokumentacji tekstowej wypada uznać za podstawowy, historyczny obowiązek polskiego językoznawstwa” (NIECKUŁA, 1990: 23). Nieckuła był przekonany o wielkim znaczeniu tego celu badawczego w wymiarze zarówno czysto poznawczym, jak i praktycznym.

5. Technika zbierania materiału i przygotowanie materiału do badań

W badaniach prowadzonych przez zespół Lubasia za jeden z istotniejszych czynników decydujących o jakości uznano naturalność mówienia. Jej gwarantem miały stać się rozmowy nagrywane z podsłuchu, gdyż jawność nagrania i obecność naukowca jako jednego z interlokutorów niweczy naturalność kontaktu. Ostatecznie zebrany materiał stanowią głównie rozmowy swobodne, ale tylko nieliczne z nich opatrzone informacją o nagrywaniu z podsłuchu, zazwyczaj też uczestniczyli w nich studenci polonistyki lub asystenci Zakładu Dialektologii IJP UŚ. Wydaje się, że z nagraniem z podsłuchu wiążą się nie tylko dostrzegane problemy prawne, ale także organizacyjne. W wyniku projektu prowadzonego w tymże Zakładzie przygotowano dwutomowy (trzyczęściowy) wybór najbardziej typowych tekstów, obejmujący 67 transkryptów pogrupowanych według miast, w których dokonywano nagrania. Teksty były spisywane (transkrypcja półfonetyczna) przez autorów nagrań, następnie przesłuchiwane i sprawdzane przez Wilkonia, po czym jeszcze raz przez Lubasia, który we wstępie do drugiego tomu wyboru tekstów przyznał, że największą trudnością było ustalanie długości pauz (LUBAŚ, 1980: 7).

W Krakowie początkowo – wzorem badań dialektologicznych – planowano wykorzystywanie kwestionariuszy fonetycznych i leksykalnych (DUNAJ, 1981: 7). Po okresie pierwszych badań zrezygnowano z ich wykorzystywania na rzecz wywiadów lub nagrań z podsłuchu. Dzięki temu uzyskano wiarygodniejszy, różnorodny materiał: „spontaniczne teksty z różnych środowisk” (DUNAJ, 1989: 15)³. Już na dość wczesnym etapie projektu uznano bowiem, że w badaniu socjolingwistycznym najcenniejsze są spontaniczne wypowiedzi utrwalone w sytuacji nieoficjalnej: w gronie rodziny, przyjaciół czy znajomych (DUNAJ, 1987: 32–33). Nagrywanie krakowian zostało rozpoczęte pod koniec 1977 roku i było prowadzone przez cały 1978 rok. W tym czasie zgromadzono około 400 godzin nagrań. Jednocześnie rozpoczęto spisywanie wypowiedzi z taśm magnetofonowych. Teksty zapisywano sławistycznym alfabetem fonetycznym, „zaznaczając w nich, w miarę możliwości, jak najskrupulatniej, pauzy, potknięcia, dźwiękowe elementy paralingwistyczne. Prowadzono odsłuchiwanie zbiorowe oraz zespołową weryfikację zapisu. W sumie nad tekstem pracował zespół 4–5 osób. Do druku przygotowano wybór najciekawszych tekstów” (DUNAJ, 1989: 15).

³ Por. BIEŃKOWSKA, UMIŃSKA-TYTOŃ (2010).

W Łodzi nagrywano dialog z eksploratorem lub dialogi między informatorami. Kamińska podkreśla, że jest to bardzo czasochłonne i trudne technicznie. Szczegółowo opisuje rolę nagrywającego rozmowę (KAMIŃSKA, 1979: 90). Wskazówki dotyczą także etapu przygotowawczego. Badaczka zaznacza, że nie nagrywanie z ukrycia, ale umiejętne poprowadzenie rozmowy przez nagrywającego decyduje o wartości zebranego materiału: „[...] przy silnym zaangażowaniu w treść wypowiedzi informator zapomina o formie, nie upiększa jej, ale używa codziennych, potocznych form” (KAMIŃSKA, 1979: 80). Do każdego nagrania dołączano charakterystykę językową informatora: jakim wariantem polszczyzny się posługuje, czy zwraca uwagę na dobór środków językowych, czy w trakcie rozmowy zmienia rejestr polszczyzny. Nagrane teksty przepisywano odręcznie z wykorzystaniem alfabetu fonetycznego. Przerwy w wypowiedzi zaznaczano pionowymi kreskami (pauza krótka – jedna kreska, pauza długa – dwie kreski. Oznaczano też imitowanie cudzego głosu czy cudzego dialogu przez informatora. Opatrywano też uwagami „ciekawe, zdaniem eksploratora, formy wyrazowe, wyrazy i zwroty frazeologiczne” (KAMIŃSKA, 1979: 90).

W ośrodku łódzkim opracowywano również kartotekę, która miała być wykorzystana do redagowania słownika. Odnotowywano w niej zleksykalizowane zjawiska fonetyczne (*krzciny*, *ochwiara*), słowotwórcze (*rozpuścicy*, *pobudynek*), leksykalne, składniowe i słowotwórcze (opis fiszki – KAMIŃSKA, 1979: 90). W kartotece znalazły się też formy nie nagrane, lecz zasłyszane u nagrywanego. Wiek mówców odnotowywany na fiszkach wyrażano w trzech grupach: do lat 30, do 60, starsi. Całość nagranych (i zasłyszanych) wypowiedzi miała służyć do opisu polszczyzny mówionej łódzian, natomiast wybór wypowiedzi miał zostać wydany w postaci antologii łódzkich tekstów. Kamińska rozważała wydanie wyboru tekstów w pisowni półfonetycznej lub nawet w zapisie ortograficznym („w pisowni literackiej”), żeby mogli z niego też skorzystać socjologowie, etnografowie, historycy.

We Wrocławiu również uznano, że najlepszym materiałem do badań byłyby nieoficjalne wypowiedzi nagrywane z ukrycia bez udziału nagrywającego. Jednocześnie podkreślono, że jednak najważniejsza w czasie nagrania jest rola eksploratora, który może wpłynąć na swobodę wypowiedzi i zadbać o naturalną atmosferę całej sytuacji. Dopuszczano możliwość uzupełniania danych (w razie stwierdzenia takiej potrzeby). Postanowiono nie nagrywać według wcześniej ustalanego klucza. Zdecydowano jedynie, że zbieranie materiałów rozpocznie się od najmłodszej generacji – dzieci ze szkół podstawowych. Uznano bowiem, że dotarcie do rozmówców będzie najłatwiejsze w zaprzyjaźnionych szkołach. Planowano, że w miarę bogacenia się zbioru kryteriów doboru mówców będą one precyzowane. Transkrybowanie miało się odbywać z zastosowaniem alfabetu fonetycznego tylko wtedy, kiedy u rozmówcy zostaną stwierdzone wyraźne cechy regionalne czy gwarowe. Przypuszczano, że w wymowie wrocławian pojawi się obficie wariacja fonetyczna w ramach dialektu kulturalnego. Miała być ona odnotowana w ogólnym opisie tekstu, a w transkryptach – zaznaczona pisownią półfonetyczną. Kierowano się ekonomizacją wysiłku oraz faktem, że kasety będą zarchiwizowane, dzięki czemu zawsze będzie można je ponownie przeanalizować. Planowano opublikowanie antologii przetran-skrybowanych tekstów.

6. Główne kierunki badań w XX wieku

Materiał zebrany przez zespół Uniwersytetu Śląskiego w Katowicach był oczywiście szansą na kontynuację badań historycznojęzykowych zainicjowanych przez Stanisława Rosponda. Jednak tak duży i różnorodny zbiór tekstów otworzył przede wszystkim możliwość badania języka mówionego w miastach śląskich i zagłębiowskich pod względem różnicowania wiekowego, zawodowego, etnicznego i kulturowego mieszkańców. Specyfika badanego obszaru stała się także zachętą do zajęcia się zagadnieniem polityki językowej w środowisku wielkoprzemysłowym (LUBAŚ, 1975).

Nowy materiał pokazał przede wszystkim, że w kontekście wypowiedzi mówionych redefinicji domaga się wypowiedzenie. Poza tym uznano, że warto podjąć zagadnienie segmentacji tekstu, typów pauz, nakładania się replik, intonacji czy elementów paralingwistycznych w wypowiedzi (LEBDA, 1979). Jeszcze jednym nowym zainicjowanym zagadnieniem badawczym stała się analiza naturalnych wypowiedzi językowych przygotowywanych do edycji (nazwano ten proces „fryzowaniem” tekstu).

W ośrodku krakowskim szybko zarysowały się zakresy badawcze rozwijane przez poszczególnych członków zespołu projektowego (DUNAJ, 1981: 8; 1984: 7–8). Były to między innymi ogólne zagadnienia socjolingwistyczne o charakterze metodologicznym (Dunaj), fonetyczne różnicowanie środowiska wielkomiejskiego (Kurek), leksyka mówionej odmiany języka (Kowalik, Frodyma, Ożóg), składnia oraz pragmatyka lingwistyczna (Awdiejew, Labocha, Ożóg), podkrakowska gwara na wiejskich terytoriach wchłoniętych przez miasto (Kąś).

Badania łódzkie zostały zaplanowane wielokierunkowo. Po pierwsze, postanowiono prowadzić analizę w formie studium przypadku (w poszukiwaniu wariantywności językowo-stylistycznej uwarunkowanej sytuacyjnie): po badaniach pilotażowych typuje się pracowników łódzkiego zakładu przemysłowego, następnie nagrywa się próbki ich rozmów w różnych sytuacjach: na stanowisku pracy, z personelem inżynierjno-technicznym, z prowadzącym badania, a także – w środowisku domowym. Po drugie, zdecydowano się również na badania pogranicza Wielkiej Łodzi z okalającymi wsiami, w których mieszka ludność łącząca pracę na roli z pracą w przemyśle. Zaplanowano również drobniejsze tematy: formacje ekspresywne łodzian różnych pokoleń, język codzienny na targowiskach miejskich i w sklepach handlu uspołecznionego – analiza wybranych form językowych; składnia wypowiedzi nagrywanych z ukrycia w warunkach domowych; wyrazy obelżywe, wyzwiska, przewziska, zdrobnienia imion; polszczyzna dzieci w wieku przedszkolnym (generacja najmłodsza).

Wrocławskie badania były skoncentrowane głównie wokół zjawisk fonetycznych (NIECKULA, 1992). Niestety, nie zrealizowano bardzo ciekawego pomysłu na badanie idiolektów. Jak już zostało wspomniane, przypuszczano, że wśród badanych nie zaobserwuje się systemowych, wspólnych dla wszystkich mieszkańców, zjawisk wyróżniających polszczyznę wrocławian. Między innymi z tego powodu Nieckula uważał, że warto koncentrować się na pojedynczych mówcach i ich rodzinach: Pisał: „[...] we Wrocławiu mamy szczególne powody, aby brać pod uwagę wielką wariantywność na poziomie idiolektu (lub tego, co można nazwać idiolektem), czyli liczyć się ze znacznym różnicowaniem osobniczym i rodzinnym” (NIECKULA, 1990: 31).

Studia nad literaturą przedmiotu wskazują, że w drugiej połowie lat 90. ubiegłego wieku badania nad polszczyzną mówioną miast w ramach opisywanego uprzednio projektu nie były kontynuowane. Jednak znalazły trwałe miejsce w historii badań socjolingwistycznych.

7. Powrót do badań nad polszczyzną mówioną Wrocławia

Wiosną 2018 roku odnaleźliśmy materiały z projektu „Polszczyzna mówiona Wrocławia”, który był prowadzony w Instytucie Filologii Polskiej w latach 90. XX wieku. Z tamtych czasów zachowało się wiele nagrań oraz transkryptów w formie maszynopisów. Nie udało się wówczas, niestety, doprowadzić do ich publikacji. Część odnalezionych materiałów była zdekompletowana, tzn. pozostały albo tylko kasety, albo tylko transkrypty. Ostatecznie wybraliśmy do utrwalenia osiem godzin nagrań, które miały pełne metadane. Cały materiał został zdigitalizowany w Pracowni Humanistyki Cyfrowej UWr przez Piotra Malaka. Do transkrypcji nagrań zastosowaliśmy program ELAN⁴ wykorzystywany również do opracowywania tekstów gromadzonych w wyszukiwarce danych konwersacyjnych Spokes, w której ostatecznie dostępne są wrocławskie materiały z lat 90.⁵

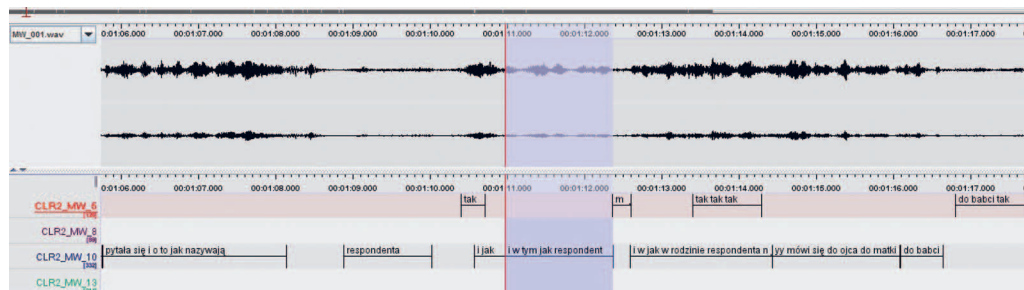
Narzędzie do transkrypcji tekstu ELAN pozwala na pracę z nagraniem tekstem, którego obraz widziany jest w postaci fali dźwiękowej z towarzyszącą jej linią czasu. Współczesny zapis transkrypcyjny utrwała się zupełnie inaczej niż w maszynopisie. Transkrypcję bowiem umieszcza się w warstwach przypisanych poszczególnym mówcom, a dodatkowa linia (*events*) pozwala na uwzględnianie zjawisk dźwiękowych towarzyszących mówieniu (np.: śmiechu – *_voice_laugh_*; kaszlu – *_voice_cough_*) oraz czasu ich trwania. Widzimy zatem nie tylko następstwo replik (jak w starych transkryptach), ale także równoległe wypowiedzi oraz sygnały z tła. Taki sposób zapisu umożliwia lepsze pokazanie dialogowości tekstu i relacji między replikami: ich następstwo, nakładanie się, pauzy między replikami mówców, wpływ zjawisk dźwiękowych towarzyszących mówieniu na jakość tekstu lub momenty rozpoczynania kolejnych replik.

Praca ze zdigitalizowanym tekstem w programie komputerowym z dodatkowym wykorzystaniem słuchawek daje, oczywiście, lepsze możliwości odsłuchu tekstu. Uzyskujemy głośniejszy, bardziej precyzyjny dźwięk, łatwość wielokrotnego odtwarzania (bez włączania i wyłączenia magnetofonu), możliwość odsłuchu pojedynczych segmentów wypowiedzi zgodnie z wybranym odcinkiem fali dźwiękowej czy linii czasu, odsłuch w ramach pętli (wielokrotnego powtarzania), odtwarzanie tekstu z uwzględnieniem różnych odcinków czasowych poprzedzających fragment lub po nim następujących. Takie ułatwienia techniczne sprawiają, że osoba transkrybująca może dokładniej zapisywać słyszane dźwięki, a mniej polegać na swojej pamięci słuchowej i przewidywaniu możliwego tekstu. Przewidywany tekst (niejako odgadywany z pamięci) zwykle bardziej odpowiada wersji ogólnej języka lub też jego odmianie pisanej. W rezultacie w przygotowanych współcześnie transkrypcjach zdecydowanie mniej jest miejsc oznaczanych jako niemożliwe do transkrypcji (ELAN:

⁴ [online: <https://tla.mpi.nl/tools/tla-tools/elan>; data dostępu: 14.02.2020].

⁵ Por. też: [online: <http://ppm.uni.wroc.pl/2018/11/22/badania-nad-polszczyzna-mowiona-wroclawia-lata-80-xx-wieku/>; data dostępu: 14.02.2020].

_gap_unclear_; w maszynopisie – /xxx/) lub urwane (ELAN:_trunc_, w maszynopisie: ..., np. pu...puść). Natomiast zdecydowanie częściej są odnotowywane pauzy wypełnione (typu yy) czy inne zjawiska charakterystyczne dla wypowiedzi spontanicznej (por. rys. 1).



Rys. 1. Fragment transkrypcji w programie ELAN

Ze względu na to, że w programie komputerowym transkrypcji zawsze towarzyszy ścieżka dźwiękowa, stosuje się w nim uproszczony zapis ortograficzny⁶ (nie jest zresztą możliwe pełne oddanie idiolektalnych cech artykulacyjnych mówców w tradycyjnie stosowanej transkrypcji półfonetycznej). Dzięki automatycznemu wyszukaniu poszczególnych realizacji danego słowa można bowiem sprawdzić sposoby jego wypowiadania w całym transkrybowanym materiale. Podobnie można porównać jakość i długość wypowiadania elementów paralingwistycznych, które są odnotowywane na przykład jako yy, ii, aa, mhm. Skoordynowanie ścieżki dźwiękowej z linią czasu i dokładny zapis kolejnych fragmentów wypowiedzi umożliwia też obserwację elementarnych jednostek wypowiedzi mówionej. Nie ma pokusy segmentacji odsłuchiwanej wypowiedzi znakami interpunkcyjnymi zgodnie z zasadami obowiązującymi w tekście pisanym. Taką właśnie metodą opracowano na nowo znaczącą część wrocławskich nagrań.

Powstały korpus jest ciekawy i niejednorodny. Tworzą go prywatne i służbowe rozmowy w różnych sytuacjach komunikacyjnych prowadzone przez dorosłych i dzieci w bardzo różnym wieku. Dorosli mówcy mają różne wykształcenie, profesje i pochodzenie regionalne: są to osoby przybyłe po wojnie do Wrocławia z różnych regionów dzisiejszej Polski, a także na przykład z Wilna, Lwowa czy okolic Wołynia.

Jak zostało to już zaznaczone, uczestnikom projektu zależało przede wszystkim na nagrywaniu spontanicznych wypowiedzi nieoficjalnych. Najpewniej z tego powodu w odnalezionych materiałach stanowią one zdecydowaną większość. Znajduje to także swoje odzwierciedlenie w nowo opracowanym korpusie. Współtworzą go głównie rozmowy towarzyskie o charakterze polilogów (*Rozmowa przy obiedzie*, *Rozmowa dwóch małżeństw*, *Rozmowy u L.*, *Aktorki*, *Kłopoty z mieszkaniem*) oraz swobodne dialogi (*Wrażenia z Ameryki*). Pewną częścią korpusu są interakcje werbalne dzieci (*Rozmowa chłopców*). Utrwalono także sposób prowadzenia rozmowy dorosłych z dziećmi (*Rozmowa z córeczką*, *Rozmowa dzieci i ich mam*, *Dzieci budują statek*). Warto podkreślić, że w części nagranych rozmów nieofi-

⁶ Uproszczony zapis ortograficzny w tym projekcie oznaczał niestosowanie wielkich liter poza nazwami własnymi.

cialnych nie bierze udziału uczestnik projektu. Takie nagrania uznawano za najcenniejsze. Analiza metadanych wskazuje, że do takich należą: *Rozmowa chłopców*, *Rozmowa dzieci i ich mam*, *Rozmowy u L.*, *Aktorki*, *Rozmowa przy obiedzie*, *Dzieci budują statek*. Natomiast *Wrażenia z Ameryki*, *Rozmowa dwóch małżeństw* i *Kłopoty z mieszkaniem* to rozmowy przeprowadzone z udziałem nagrywającego. Jednak ogląd odręcznych notatek na starych, papierowych transkryptach wskazuje, że niektóre z tych rozmów eksplorator mógł nagrywać bez wiedzy pozostałych osób (na marginesie transkryptu *Kłopoty z mieszkaniem* widnieje m.in. uwaga: „z podsłuchu”). Natomiast na transkrypcie dialogu aktorki z nagrywającym *Wrażenia z Ameryki* widnieje zapisek innego rodzaju: „swobodna rozmowa z aktorką o teatrze i wrażeniach z pobytu w Stanach Zjednoczonych”. W opisie nieoficjalnej interakcji werbalnej wyeksponowano więc cechę swobodności. Uczyniono to pewnie dlatego, by nie wątpił w nią żaden badacz, który mocno przywiązywał wagę do utrwalania wypowiedzi wyłącznie tych osób, które nie wiedzą, że są nagrywane.

Dwie rozmowy, o których nie wspomniano uprzednio, zbliżają się w swej istocie do wywiadu narracyjnego (*Pan Jan z Wilna*, *Rozmowa dorosłej córki*⁷). Są próbą uzyskania informacji o historii życia ludzi zaawansowanych wiekiem przez dużo młodsze od nich osoby. Z całą pewnością nie są to rozmowy formalne, ale różnią się co do charakteru od przywołanych wcześniej nieformalnych wypowiedzi o charakterze towarzyskim.

Jak już zostało zauważone, twórcy projektu „Polszczyzna mówiona miast” bardzo dbali o zróżnicowanie sytuacyjne nagrywanych interakcji werbalnych. Znajduje to także odzwierciedlenie w materiale wrocławskim. Szczęśliwie się więc składa, że wśród odnalezionych i nowo opracowanych nagrań ujawniono także dwa polilogi utrwalone w sytuacji formalnej. Są to *Dyskusja naukowa* i *Zebranie zakładu*.

W *Dyskusji naukowej* bierze udział 7 osób. Ich wypowiedzi to zarówno wystąpienia mówione w odmianie opracowanej (powitanie uczestników konferencji, przemówienie rektora), jak i głosy członków audytorium (np. nauczycieli), odpowiedzi referentów na pytania z sali czy informacje przekazywane uczestnikom konferencji przez organizatorów. Nagrania te pokazują, jak bardzo zróżnicowane pod względem stylistycznym mogą być wypowiedzi formułowane w sytuacji oficjalnej. Bardzo ciekawa jest również obserwacja wnikania elementów dialogu do monologowych z założenia wypowiedzi konferencyjnych. Wraz z nagraniami przygotowywanymi na potrzeby projektu GeWiss (w trakcie jego realizacji utrwalano m.in. wygłaszane na konferencjach referaty z dyskusją po ich wygłoszeniu) mogą chociażby stanowić element porównawczy w kolejnych badaniach z zakresu współczesnej polszczyzny naukowej w odmianie mówionej.

Na tle całości nowo opracowanego wrocławskiego zbioru wyjątkowym nagraniem jest *Zebranie zakładu* (zachowało się z pełnymi metadanych i maszynopisem starego transkryptu). Jest to rozmowa 15 naukowców podczas roboczego spotkania w IFP UW r tuż przed przystąpieniem do badań miejskiej polszczyzny mówionej. W trakcie tej rozmowy formułowane są główne cele projektu w ramach ogólnopolskiego programu węzłowego. Biorą w niej udział naukowcy różnych pokoleń: zarówno doświadczeni profesorowie, jak

⁷ *Rozmowa dorosłej córki* została ujawniona w trakcie transkrybowania ok. połowy trzydziestominutowego pliku zatytułowanego (na wzór tytułu odpowiedniej kasety magnetofonowej) *Pan Jan z Wilna*. Nie ma do niej żadnych metadanych.

i adiunkci ostrożnie formułujący tematy swoich prac habilitacyjnych (niektóre z nich zostały później faktycznie zrealizowane). Już sam odsłuch nagrania *Zebranie zakładu* przekonuje do idei utrwalania rozmów z przedstawicielami nauki choćby w formie wywiadów naracyjnych. Mają one bowiem wielowymiarową wartość, trudną do przecenienia nie tylko w kontekście badań socjolingwistycznych.

Trzeba podkreślić, że przy ponownym opracowywaniu starych nagrań intensywnie pracował duży zespół pracowników różnych ośrodków naukowych. W dalszej części artykułu zostaną wymienione wszystkie osoby, które wzięły udział w opisywanym projekcie.

Tabela 1

Projekt „Mowa_Wrocław” (maj-czerwiec 2018)

Etapy projektu	Uczestnicy projektu
Inwentaryzacja kaset i transkryptów, wstępne opracowanie metadanych	Monika Zaško-Zielińska UWr, Anna Majewska-Tworek UWr
Współpraca z CLARIN-em	Piotr Pęczik UŁ
Szkolenie w transkrypcji (ELAN)	Piotr Pęczik UŁ, Maciej Grabski UŁ, Kamil Wabnic UWr
Dygitalizacja nagrań	Piotr Malak UWr
Poprawianie jakości nagrań po dygitalizacji	Kamil Wabnic UWr
Przetworzenie metadanych dla CLARIN-u	Maciej Grabski UŁ
Anotacja morfosyntaktyczna i zindeksowanie danych w systemie Spokes	Piotr Pęczik UŁ, Paweł Kowalczyk UŁ
Anotacja czasowa narzędziami mowa.clarin.pl	Michał Kowalczyk
Przetranskrybowane nagrania (symbol, tytuł) i autorzy transkrypcji	MW_001: <i>Zebranie zakładu</i> – Anna Majewska-Tworek (IFP UWr) MW_002: <i>Rozmowy z córeczką</i> – Paulina Witkowska (IFP UWr) MW_003: <i>Rozmowa przy obiedzie</i> – Marta Śleziak (IFP UWr) MW_004: <i>Rozmowy u L.</i> – Aleksandra Płochocka (IFG UWr) MW_005: <i>Dzieci budują statek</i> – Ewa Kaczmarz (IFP UWr) MW_006: <i>Rozmowa dzieci i ich mam</i> – Kinga Sudoł (IFP UWr) MW_007: <i>Rozmowa dwóch małżeństw</i> – Kamil Wabnic (IFP UWr) MW_014: <i>Dyskusja naukowa</i> – Monika Zaško-Zielińska (IFP UWr) MW_015: <i>Kłopoty z mieszkaniem</i> – Maria Rudnicka (IFP UWr) MW_017: <i>Wrażenia z Ameryki</i> – Karolina Okurowska (IFP UWr) MW_018: <i>Pan Jan z Wilna</i> – Michalina Jaworska (IFP UWr) MW_022: <i>Aktorki</i> – Agnieszka Skwarek (IFG UWr) MW_050: <i>Rozmowa chłopców</i> – Natalia Śmiełowska (IFP UWr) MW_057: <i>Rozmowa dorosłej córki</i> – Michalina Jaworska (IFP UWr)

8. Włączenie danych do systemu Spokes

W ramach projektu CLARIN-PL od 2015 roku rozwijane i utrzymywane są dwie wersje systemu wyszukiwawczego Spokes dla korpusów polszczyzny mówionej. Pierwsza z nich pozwala na przeszukiwanie danych konwersacyjnych, które pierwotnie weszły do korpusu PELCRA (WALIŃSKI, PĘZIK, 2007) i NKJP (PĘZIK, 2012). Wyszukiwarka obsługuje specjalną składnię zapytań korpusowych (PĘZIK, 2015), a także eksportowanie wyników oraz pełnych transkrypcji zindeksowanych rozmów. Odtwarzanie fragmentów nagrań pasujących do zapytań korpusowych jest możliwe dzięki ręcznej transkrypcji rozmów oraz oznaczeniu czasu trwania poszczególnych wypowiedzi, czyli początku i końca ich wystąpienia na osi czasu. Ręczna transkrypcja rozmów jest konieczna ze względu na zaszumienie i różnorodne ekstralingwistyczne zjawiska występujące w większości nagrań, a co za tym idzie – niską skuteczność użycia systemów automatycznej transkrypcji mowy. Jednocześnie, oznaczenie granic ręcznie transkrybowanych wypowiedzi pozwala na użycie narzędzi służących przetwarzaniu mowy opracowanych w projekcie CLARIN-PL (KORŻINEK et al., 2017) do oznaczania czasu trwania słów i głosek w nagraniach. Z tej anotacji korzysta między innymi nowsza wersja wyszukiwarki o nazwie SpokesMix, która umożliwi wizualizację i eksport nie tylko wyrazów, ale ich fonemicznych segmentów w wynikach wyszukiwania (PĘZIK, 2018).

Proces włączania transkrypcji i nagrań zbioru *Mowa Wrocławia* do indeksu systemu Spokes przebiegał zgodnie z zasygnalizowaną powyżej procedurą dygitalizacji danych konwersacyjnych. Transkrypcje w formacie ELAN (.eaf) zostały najpierw podzielone na segmenty wyrazowe, a następnie poddane analizie morfosyntaktycznej za pomocą tagera APT_PL (PĘZIK, LASKOWSKI, 2017). W kolejnym kroku wyliczone zostały automatycznie iloczasy słów i głosek wewnątrz anotowanych ręcznie granic wypowiedzi. Następnie transkrypcje z anotacją czasową i lingwistyczną zostały zindeksowane w systemach Spokes i SpokesMix. Rysunek 2 ukazuje przykład zapytania o przymiotniki występujące przed rzeczownikami *sprawa*, które daje wystąpienia fraz takich, jak *ciężka sprawa*, *moja sprawa* itp. Zakres wyszukiwania jest w tym wypadku ograniczony do transkrypcji ze zbioru *Mowa Wrocławia* poprzez zdefiniowanie filtra *Source = UWR_MMW* w zaawansowanych opcjach wyszukiwania.

Transkrypcje nagrań ze zbioru *Mowa Wrocławia* włączone do indeksu SpokesMix zawierają 64 861 segmentów wyrazowych, co stanowi niespełna 2,6% wszystkich polskich danych zindeksowanych w tym korpusie. Mimo pozornie niewielkiego udziału *Mowy Wrocławia* w całej puli danych jej transkrypcje stanowią cenne uzupełnienie całości zasobów co najmniej z dwóch powodów. Po pierwsze, rejestr językowy transkrypcji oraz metodologia ich pozyskania są spójne z próbkami polszczyzny konwersacyjnej, które są głównym i unikatowym w wolnym dostępie komponentem korpusów PELCRA i NKJP udostępnianych przez system Spokes. Po drugie, utrwalone na nagraniach rozmowy pochodzą z lat 80. ubiegłego wieku, podczas gdy ogromna większość nagrań dostępnych do tej pory w zasobach Spokes pochodzi z lat 2000–2018. Dzięki wsparciu zespołu i infrastruktury CLARIN-PL nagrania sprzed około 35 lat mogą być przeszukiwane i odsłuchiwane w swobodnym dostępie przez współczesnych badaczy spontanicznej polszczyzny konwersacyjnej.

The screenshot shows the Spokes search interface. At the top, a search box contains the query '<tag=adj.*> sprawa**'. Below it, the 'Dismax' section shows 'source_corpus:UWR_MMW'. A filter section includes 'NOT', 'AND', and 'OR' options, with 'Source' set to 'equal' and 'UWR_MMW'. A 'Query name' field and a 'Mark as default' checkbox are also visible. Below the search interface, the 'Concordances' section is active, showing a table with 37 entries. The table has columns for '#', 'Left', 'Match', 'Right', 'Wav', 'Emu', and 'Audio'. The 'Match' column highlights the word 'sprawa' in various contexts from the corpus.

#	Left	Match	Right	Wav	Emu	Audio
1	no	ciężka sprawa	.			
2	. bo ja siedzę a to nie jest mieszkanie moje tylko ziemia . tu jest	ciężka sprawa	z tym ja siedzę bo siedzę .			
3		jest to konkretna spr				
4	owę więc wiesz _gap_unclear_ wiesz ja uważała m że to nie jest	moja sprawa	_gap_unclear_ wiesz			
5	sprawa nauczycieli tutaj yy była i pani o tym mówiła a ja yy	najważniejsza	: to jest w ogóle to			
6	yy to była by chyba yy ta	najważniejsza	:			
7	im się że to przedstawienie opowiada o naszych problemach o	naszycy sprawy	okazało się że jest ono bardziej uniwersalne niż nam się wyd...			
8	no to jest kwestia to sprawa scenografii ale w tekście jest po	prostu sprawa	prosta ze kilku mężczyzn buduje mur tak ?			
9	nie to są dwie	różne sprawy	odbiór za granicą i granie klasyki wydaje mi się że			

Rys. 2. Wyniki zapytania o przymiotniki przed rzeczownikiem *sprawa* w Spokes (zbiór: *Mowa Wrocławia*)

Literatura

- AWDIEJEW A., 1981: *Składnia języka mówionego jako proces bezpośredniego wyboru leksykalnego i formalnego*. W: DUNAJ B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 93–102.
- BIEŃKOWSKA D., UMIŃSKA-TYTOŃ E., 2010: *Regionalizacja czy unifikacja leksyki? Uwagi na marginesie kwestionariusza do badań regionalizmów leksykalnych*. „Studia Językoznawcze” IX, s. 23–31.
- BRUGMAN H., RUSSEL A., 2004: *Annotating Multi-media / Multi-modal Resources with ELAN*. In: LINO M., XAVIER M., FERREIRA F., COSTA R., SILVA R., eds.: *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*. Paris, s. 2065–2068.
- DUNAJ B., 1981: *Badania języka mówionego mieszkańców Krakowa*. W: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 7–10.
- DUNAJ B., 1984: *Dalsze badania języka mówionego mieszkańców Krakowa*. W: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 7–11.
- DUNAJ B., 1987: *O kilku problemach metodologicznych socjolingwistycznych badań języka mówionego (w świetle dotychczasowych doświadczeń)*. „Socjolingwistyka” VI, s. 29–37.
- DUNAJ B., 1989: *Język mieszkańców Krakowa*. Cz. 1: *Zagadnienia teoretyczne, fonetyka, fleksja*. Kraków.
- DUNAJ i in., 1979: DUNAJ B., AWDIEJEW A., KOWALIK J., KUREK H., OŻÓG K.: *Ogólne założenia badań języka mówionego Krakowa i Nowej Huty*. „Socjolingwistyka” II, s. 77–85.

- FRODYMA M., 1981: *W sprawie słownika polszczyzny mówionej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 89–92.
- FURDAL A., 1966: *Podział polskich dialektów miejskich*. „Rozprawy Komisji Językowej WTN” VI, s. 155–161.
- FURDAL A., 1973: *Materiały do kwestionariusza do badania dialektów kulturalnych i miejskich w Polsce*. „Rozprawy Komisji Językowej WTN” IX, s. 41–59.
- KAMIŃSKA M., 1979: *Założenia metodyczne badania polszczyzny mówionej w Łodzi*. „Socjolingwistyka” II, s. 77–94.
- KĄŚ J., 1984: *Problem wariantowości fonetycznej w gwarach ludowych w obrębie Krakowa (na przykładzie a pochylonego)*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 73–91.
- KORŻINEK i in., 2017: KORŻINEK D., MARASEK K., BROCKI Ł., WOŁEK K.: *Polish Read Speech Corpus for Speech Tools and Services*. In: *Selected Papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016*. Linköping, s. 54–62.
- KOWALIK J., 1981: *Niektóre regionalizmy leksykalne w mowie młodzieży szkolnej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 69–87.
- KUREK H., 1981: *Próba zastosowania metod statystycznych do badania fonetyki języka mówionego mieszkańców Krakowa. Na przykładzie wygłosowego -q*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 21–33.
- KUREK H., 1991: *Zróźnicowanie fonetyczne języka mówionego mieszkańców Krakowa (wybrane zagadnienia)*. W: DUNA J. B., OZÓG K., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 3. Kraków, s. 147–155.
- LABOCHA J., 1981: *Przerwania jako sygnały trudności wyboru leksykalnego i formalnego w polszczyźnie mówionej*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 1. Kraków, s. 111–120.
- LEBDA R., 1979: *Dźwiękowe elementy parajęzykowe. Na przykładzie mówionego języka robotników województwa katowickiego*. „Socjolingwistyka” II, s. 149–160.
- LUBAŚ W., 1975: *Badania nad językiem jako punkt wyjścia dla ustalania i realizowania polityki językowej*. W: KANTYKA J., LUBAŚ W., red.: *Rola nauk społecznych w humanizacji środowiska wielkoprzemysłowego*. Katowice, s. 62–69.
- LUBAŚ W., 1978: *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia*. T. 1. Katowice.
- LUBAŚ W., 1980: *Teksty języka mówionego mieszkańców miast Górnego Śląska i Zagłębia*. T. 2. Katowice.
- NIECKULA F., 1990: *Uwagi ogólne do badań nad polszczyzną Wrocławia*. W: NIECKULA F., red.: *Polszczyzna mówiona Wrocławia*. Cz. 1. Wrocław, s. 8–38.
- OZÓG K., 1984: *Grzecznościowe akty mowy*. W: DUNA J. B., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 2. Kraków, s. 147–157.
- OZÓG K., 1991: *Jednostki otwierające i zamykające replikę w dialogu*. W: DUNA J. B., OZÓG K., red.: *Studia nad polszczyzną mówioną Krakowa*. [T.] 3. Kraków, s. 71–89.
- PĘZIK P., 2018: *Increasing the Accessibility of Time-Aligned Speech Corpora with Spokes Mix*. European Language Resources Association (ELRA). Miyazakis, s. 4297–4300.
- PĘZIK P., 2012: *Język mówiony w NKJP*. W: PRZEPIÓRKOWSKI A., BAŃKO M., GÓRSKI R., LEWANDOWSKA-TOMASZCZYK B., red.: *Narodowy Korpus Języka Polskiego*. Warszawa, s. 37–47.
- PĘZIK P., 2015: *Spokes – a Search and Exploration Service for Conversational Corpus Data*. In: *Selected Papers from the CLARIN 2014 Conference, October 24–25, 2014, Soesterberg, The Netherlands*. Linköping, s. 99–109.

- PĘZIK P., LASKOWSKI S., 2017: *Evaluating an Averaged Perceptron Morphosyntactic Tagger for Polish*. In: *Human Language Technologies as a Challenge for Computer Science and Linguistics. 8th Language & Technology Conference November 17–19, 2017, Poznań, Poland. Proceedings*. Poznań, s. 372–376.
- WALIŃSKI J., PĘZIK P., 2007: *Web Access Interface to the PELCRA Referential Corpus of Polish*. In: WALIŃSKI J., KREDENS K., GOŹDŹ-ROSKOWSKI S., Eds.: *Corpora and ICT in Language Studies. “Łódź Studies in Language” XIII*. Ed. B. LEWANDOWSKA-TOMASZCZYK. Frankfurt am Main–Berlin–Bern–Bruxelles–New York–Oxford–Wien, s. 65–86.
- WILKOŃ A., 2000: *Typologia odmian językowych współczesnej polszczyzny*. Katowice.