

Ilona Kulak
Wydział Polonistyki
Uniwersytet Jagielloński w Krakowie
e-mail: ilonakulak21@gmail.com

Korpus spiskich tekstów i nagrań gwarowych. Założenia projektowanej bazy i możliwości jej wykorzystania jako źródła badań interdyscyplinarnych

Abstract

The Electronic Corpus of Texts and Recordings from Poland's Spisz Region. The Assumptions of The Database and The Possibilities of its Use as a Source of Interdisciplinary Research

The article presents the corpus of texts and recordings from Poland's Spisz region as a source of material for the study of culture and history of this historical, geographical and ethnographic land located in the Western Carpathians. The paper deals with electronic database of dialectal transcripts and recordings collected during research within the project *The language of the Spisz region inhabitants*. *The corpus of texts and dialect recordings* carried out by The Institute of the Polish Language at the Polish Academy of Sciences. The main aims of the research project is to document spoken language of inhabitants Spisz region and to develop an electronic corpus of dialect texts. Nevertheless rich and various material could be useful not only in language studies, but also in historical, ethnographic and folklore studies. Therefore the article contains the general information about the project such as exploratory research or transcribing texts and examples of informants' statements.

Keywords: electronic database, database online, corpus, Poland's Spisz region, spoken language, dialect texts.

Wprowadzenie

Bezwzględna „potrzeba wykorzystania nowych koncepcji i narzędzi badawczych, które pozwoliłyby przetworzyć i zrozumieć ogromne ilości informacji” stała się jednym z priorytetów współczesnych nauk humanistycznych (Radomski, Bomba 2013: 7). Tradycyjne metody badawcze, a także sposoby archiwizacji przestały być wystarczające. Otwarta forma publikowania wyników badań czy rozpowszechniania zgromadzonych materiałów niweluje ograniczenia związane z tradycyjnymi formami drukowanymi, a tym samym zwiększa szanse dotarcia do szerszego grona zainteresowanych daną tematyką. Tendencja ta żywa jest szczególnie w językoznawstwie komputerowym, zwłaszcza korpusowym, zajmującym się między innymi magazynowaniem i przetwarzaniem cyfrowych zbiorów tekstów (por. Lewandowska-Tomaszczyk 2005: 11–12). Oprogramowania umożliwiające stworzenie elektronicznej bazy transkrypcji i jej przeszukiwanie stają się współcześnie trafnym sposobem gromadzenia oraz archiwizowania materiałów dialektalnych – źródła wiedzy o języku i kulturze ludowej.

Zachodzące od kilkudziesięciu lat gwałtowne przeobrażenia społeczne „prowadzą szybkie i nieodwracalne zmiany w rodzimych systemach gwarowych” oraz w sposobie komunikowania się wspólnoty wiejskiej (Kurek 1995: 147; por. Gajda 2006: 15). Gwary, podobnie jak każdy język, nie są monolitami, lecz ewoluują w czasie i w przestrzeni przez ich wewnętrzny rozwój (por. Kurek 2006: 53). Na tempo i zakres przeobrażeń językowych społeczeństwa wpływają czynniki pozajęzykowe: przemiany społeczne, demograficzne, ekonomiczne, polityczne oraz kulturowe. Szczególne znaczenie miały zmiany, które dokonały się po drugiej wojnie światowej, między innymi migracja ludności wiejskiej do miast, podejmowanie przez mieszkańców wsi pracy w miastach, upowszechnienie szkolnictwa oraz mediów (Kurek 2006: 54; Kurek 2005: 591; por. Pelcowa 2002: 383; Grochola-Szczepanek 2013: 43–44; Kaś 2001: 191–192). Obecnie, ze względu na nasilanie się tendencji globalizacyjnych, gwara jako język mówiony jest „szczególnie narażona na zanik i bezpowrotne zapomnienie” (Grochola-Szczepanek 2006b: 167), dlatego też pilną potrzebą staje się utrwalanie tego niematerialnego dziedzictwa kulturowego społeczności wiejskich (por. Pelcowa 2012: 208), m.in. poprzez archiwizację tekstów gwarowych.

Celem artykułu jest przybliżenie kwestii związanych z pracą nad elektroniczną bazą tekstów i nagrań języka mieszkańców polskiego Spisza¹. Omówione zostaną założenia projektu, sposób pozyskiwania materiału dialektalnego, cechy

¹ Spisz – region położony w Karpatach Zachodnich, na północnym zachodzie sąsiadujący z Podhalem, na wschodzie z Szaryszem, na południu – Abovem oraz Gemerem, na południowym zachodzie – Liptowem (regiony słowackie); zajmuje ok. 3700 km². Obecnie większość obszaru należy do Słowacji, polska część obejmuje jedynie ok. 195,5 km². Polski Spisz pod względem geomorfologicznym stanowi wschodnią część Podhala; zajmuje północny fragment Zamagurza Spiskiego; jego geograficzną granicę od zachodu wyznacza Białka (oddzielająca go od Podhala), od południa

powstającego zbioru, transkrypcja tekstów, narzędzia informatyczne oraz interfejs korpusu i sposoby wyszukiwania. Szkic dostarczy także przykładów ilustrujących możliwości projektowanej bazy jako źródła wiedzy historycznej, etnograficznej i folklorystycznej. Teksty i nagrania gwarowe, obrazujące kulturę ludową, tradycje, zwyczaje, życie społeczne, mogą bowiem stanowić podstawę badań nie tylko językoznawczych, lecz także wybranych nauk społecznych.

Cele i zadania projektu

Projekt *Język mieszkańców Spisza. Korpus tekstów i nagrań gwarowych* realizowany jest w Instytucie Języka Polskiego PAN w Krakowie². Głównym jego celem jest dokumentacja języka mieszkańców polskiej części Spisza oraz opracowanie elektronicznej bazy tekstów i nagrań gwarowych, dostępnej nie tylko dla wąskiego grona specjalistów, zwłaszcza dialektologów czy etnolingwistów, lecz także dla reprezentantów innych dyscyplin naukowych, a nawet samych górali spiskich (por. Grochola-Szczepanek 2017: 107–108). Korpus języka mieszkańców Spisza przyniesie wiedzę zarówno na temat stopnia zachowania samej gwary, jej cech fonetycznych, morfologicznych, składniowych, leksykalnych, jak i kultury czy historii regionu na początku XXI wieku³.

„Projekt będzie jednym z pierwszych (o ile w ogóle nie pierwszym) korpusów całościowo przedstawiających język mówiony konkretnego regionu w Polsce online” (Grochola-Szczepanek 2017: 107). Obecnie największym internetowym opracowaniem dialektologicznym jest kompendium *Dialekty i gwary polskie* pod redakcją Haliny Karaś (2010), w którym, oprócz informacji o dialektach, pojawiają się nagrania tekstów gwarowych różnych regionów. Zdigitalizowane wywiady dialektologiczne (prawie dwa tysiące godzin nagrań) z mapami, opisami dialektów i gwar polskich udostępnia także witryna internetowa Dialektologiczna Fonoteka Wydziału Polonistyki UJ (2014–2016). Dotychczas jednak nie utworzono bazy poświęconej kompleksowemu opracowaniu określonej gwary czy – szerzej – dialektu polskiego. Powstał co prawda elektroniczny korpusu tekstów polskiej gwary wsi Maćkowce na Podolu (Krawczyk-Wieczorek 2012), jednak nie został

i południowego wschodu – Słowacja, natomiast od północy i północnego wschodu – Dunajec (Biały 1987: 7–8).

² Projekt finansowany jest przez Narodowy Program Rozwoju Humanistyki (nr ref. 1bH 15 0166 83), kierownikiem projektu jest Helena Grochola-Szczepanek. Prace trwają od 2015 r., zakończenie projektu zaplanowane jest na koniec 2018 r. Obecnie trwają zaawansowane prace m.in. nad ujednocnieniem sposobu zapisów, transkrypcją ostatnich nagrań, lematyzacją form gwarowych oraz testowaniem całej bazy.

³ Dotychczasowe badania nad językiem mieszkańców Spisza oraz przegląd publikacji na temat gwary regionu (atlasy, teksty gwarowe, opracowania językoznawcze) przedstawia szczegółowo H. Grochola-Szczepanek (2012: 45–51).



Rycina 1. Podział administracyjny Spisza. Miejscowości, w których prowadzono badania terenowe

Źródło: *Korpus Spiski. Wersja demonstracyjna*, <https://spisz.ijp.pan.pl/Calosc> (dostęp: 15.05.2018).

udostępniony w serwisie internetowym (Karaś 2014: 39). Działania tego typu są natomiast obecnie rozpowszechnione w badaniach światowych (por. Grochola-Szczepanek 2017: 107).

Materiał językowy do korpusu spiskich tekstów pozyskano podczas wywiadów z mieszkańcami wszystkich piętnastu wsi polskiego Spisza⁴: Czarnej Góry, Dursztyna, Falsztyna, Frydmana, Jurgowa, Kacwina, Krempach, Łapszanki, Łapsz Niżnych, Łapsz Wyżnych, Niedzicy, Niedzicy-Zamek⁵, Nowej Białej, Rzepisk oraz Trybsza (zob. rycina 1). Miejscowości te leżą w małopolskim obszarze dialektal-

⁴ Większość miejscowości polskiego Spisza należy do gminy Łapsze Niżne (powiat nowotarski), Dursztyn, Krempachy i Nowa Biała położone są w obrębie gminy Nowy Targ (powiat nowotarski), Czarna Góra, Jurgów i Rzepiska włączone są natomiast do gminy Bukowina Tatrzańska (powiat tatrzański).

⁵ Niedzica-Zamek do roku 2014 stanowiła przysiółek Niedzicy noszący nazwę Zamek. Obecnie jest jedną z dziewięciu miejscowości należących do gminy Łapsze Niżne. Zob. *Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 13 grudnia 2013 r. w sprawie ustalenia, zmiany i zniesienia urzędowych nazw niektórych miejscowości oraz ustalenia nazwy obiektu fizjograficznego*, Dziennik

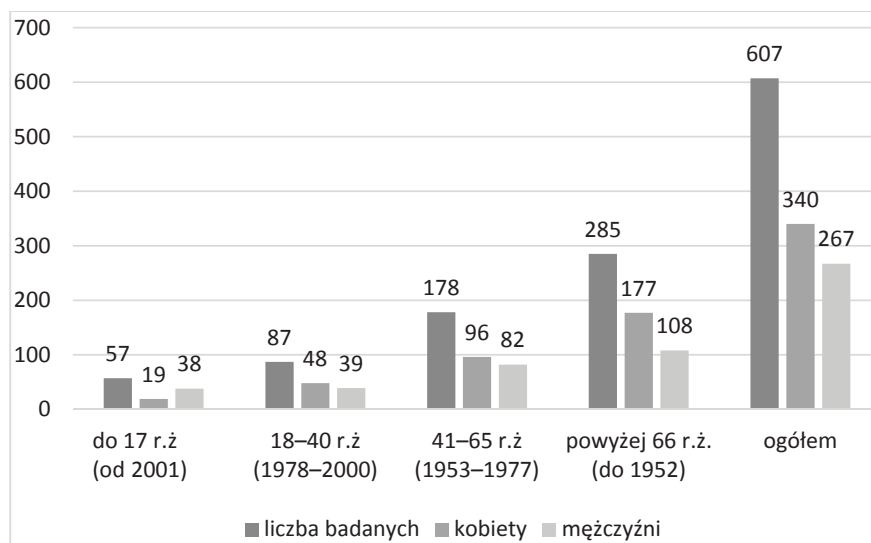
nym, a charakterystyczna dla tego regionu jest gwara spiska. Język mieszkańców tych wsi, które są położone w południowo-zachodniej części regionu (Rzepiska, Jurgów, Czarna Góra), zbliżony jest do gwary podhalańskiej. Realizacja badań korpusowych w polskiej części Spisza wynika między innymi z czynników geograficznych i administracyjnych, a także możliwości zespołowych i finansowych projektu. W przyszłości badania mogą być prowadzone również w spiskich miejscowościach na terenie Słowacji. Elektroniczną bazę tekstów i nagrań można bowiem w dowolnym momencie edytować oraz poszerzać o nowe materiały.

Sposób prowadzenia badań

Wywiady przeprowadzono z respondentami w różnym wieku, zarówno starszymi, jak i młodszymi. Często w ich pozyskiwaniu zaangażowani byli między innymi mieszkańcy zajmujący się lokalną działalnością kulturalną oraz eksploratorzy-autochtoni. Powadzono rozmowy indywidualne oraz grupowe, zwykle tak zwane diady, czyli wywiady z parą respondentów, najczęściej małżeństwem (Grochola-Szczepanek 2006a: 25). Sporadycznie zdarzały się badania fokusowe, przeprowadzane w grupie rodzinnej (dziadkowie, rodzice, dzieci) bądź sąsiedzkiej (grupa mężczyzn lub kobiet w tym samym wieku; grupa mieszana pod względem wieku i płci) (Grochola-Szczepanek 2017: 110). Metoda ta pozwala na zgromadzenie bogatego i różnorodnego materiału w stosunkowo krótkim czasie oraz warunkach sprzyjających bardziej swobodnym wypowiedziom informatorów (Grochola-Szczepanek 2006a: 34). W wywiadach grupowych dochodzi jednak wielokrotnie do nakładania się głosów respondentów, co wpływa niekorzystnie na jakość nagrania oraz nastęrcza trudności podczas transkrypcji (por. Grochola-Szczepanek 2017: 110). Tematyka rozmów była bliska mieszkańcom, ukierunkowywały ją najczęściej wiek, doświadczenie czy zainteresowania informatora. Pomoc w rozmowach z informatorami stanowił kwestionariusz opracowany na potrzeby badań, składający się z ponad 250 pytań skupionych wokół następujących tematów: szkoła i praca (na gospodarstwie, poza gospodarstwem), dom, rodzina, małżeństwo, model życia, wieś, region, mieszkańcy, współczesność, nowoczesność, kobieta, mężczyzna. Schemat wywiadu zawierał także pytania dotyczące samego respondenta. Pozyskane metadane, na przykład wykształcenie, zawód, pochodzenie czy dłuższe pobyty poza wsią, pozwalają wytłumaczyć zachowania językowe informatora, między innymi utrzymującą się cechę fonetyczną czy leksykalną charakterystyczną dla innej gwary, dominujący kod ogólny lub mieszany.

Przeprowadzono wywiady z ponad sześćsiuset informatorami ze wszystkich piętnastu miejscowości spiskich (zob. rycina 2, tabela 1). Na jedną wieś przy-

padło około 25–40 wywiadów, liczących średnio około 40 minut. Uzyskano tym samym ponad 400 godzin nagrań. Liczba plików dźwiękowych zamieszczonych w bazie będzie jednak mniejsza niż liczba nagranych rozmów. Jak zakładał projekt, w korpusie dostępnych będzie około 250 godzin nagrań kwalifikujących się pod względem jakościowym do transkrypcji (Grochola-Szczepanek 2017: 110). Mimo prób równoważenia materiału pod względem wieku, płci oraz miejsca zamieszkania informatorów przeważać będą wywiady z najstarszymi mieszkańcami wsi spiskich. W tym momencie bowiem najistotniejsza jest archiwizacja mowy użytkowników posługujących się wyłącznie kodem gwarowym, dlatego przewaga respondentów powyżej 65. roku życia wydaje się zasadna mimo potencjalnych trudności w opracowywaniu zestawień statystycznych przez przyszłych badaczy. Zaznaczyć należy, że dysproporcja liczby badanych w poszczególnych grupach wiekowych oraz wsiach wynika z wielu czynników, na przykład ogólnej liczby mieszkańców danej miejscowości, pracy w mieście lub za granicą, przewagi liczebnej kobiet w wieku poprodukcyjnym, niechęci informatorów (częściej mężczyzn niż kobiet), a nawet terminu badań terenowych. Istotna jest także zależność długości wywiadów od wieku respondentów. Z osobami starszymi wywiady są zwykle długie, wielowątkowe, mają rozbudowaną warstwę wypowiedzi informatora. Wywiady z młodszymi mieszkańcami są natomiast krótsze, z dużą liczbą pytań eksploratora i niezbyt rozbudowaną warstwą wypowiedzi informatora.



Rycina 2. Liczba respondentów z uwzględnieniem płci i wieku

Źródło: opracowanie własne

Tabela 1. Liczba informatorów z poszczególnych miejscowości z uwzględnieniem wieku i płci (K = kobieta, M = mężczyzna)

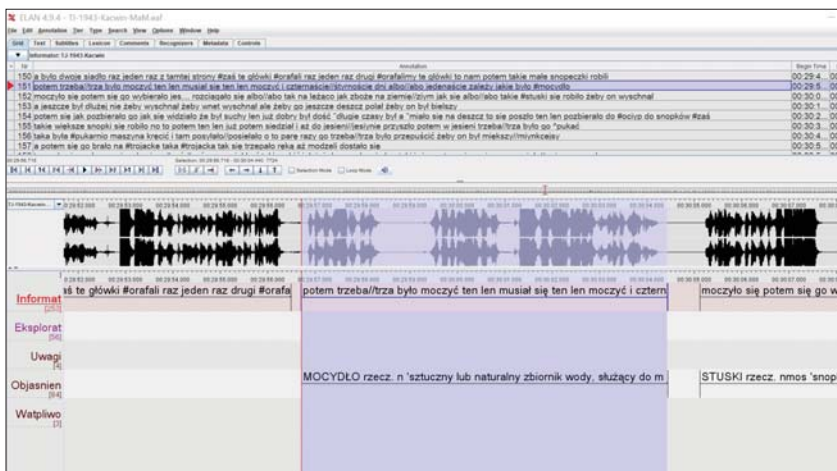
	do 17 r.ż. (od 2001)		18–40 r.ż. (1978–2000)		41–65 r.ż. (1953–1977)		powyżej 66 r.ż. (do 1952)	
	K	M	K	M	K	M	K	M
<i>Czarna Góra</i>	1	3	4	3	6	6	15	11
<i>Dursztyn</i>	0	3	2	2	8	3	13	7
<i>Falsztyn</i>	1	0	3	1	6	3	6	3
<i>Frydman</i>	2	2	0	2	6	7	17	7
<i>Jurgów</i>	0	0	6	2	6	4	9	6
<i>Kacwin</i>	2	6	2	2	14	7	20	12
<i>Krempachy</i>	2	2	6	4	5	8	13	4
<i>Łapszanka</i>	1	6	3	2	5	5	8	7
<i>Łapsze Niżne</i>	1	1	4	2	4	3	11	8
<i>Łapsze Wyżne</i>	0	3	1	0	4	2	11	6
<i>Niedzica</i>	3	0	5	6	6	6	15	8
<i>Niedzica-Zamek</i>	0	0	4	1	3	2	4	5
<i>Nowa Biała</i>	1	2	3	4	9	8	9	5
<i>Rzepiska</i>	2	2	3	6	5	11	14	9
<i>Trybsz</i>	3	8	2	2	9	7	12	10

Źródło: opracowanie własne.

Sposób transkrypcji zgromadzonych tekstów i nagrań

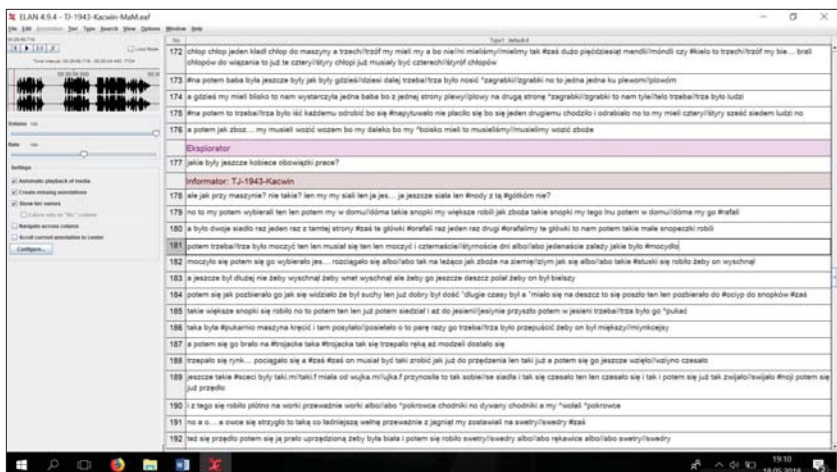
Zarejestrowane przy pomocy dyktafonu rozmowy transkrybowane są w oprogramowaniu ELAN⁶, służącym głównie do archiwizacji języka. Każdy plik dzielony

⁶ Oprogramowanie ELAN utworzone zostało przez zespół Archiwum Językowego Instytutu Psycholingwistyki im. Maxa Plancka w Holandii. Program służy do ręcznego tworzenia nieograniczonej liczby złożonych adnotacji w zasobach audio i wideo. Narzędzie to umożliwia połączenie zapisu z dźwiękiem, dzięki czemu można jednocześnie widzieć zapisaną wypowiedź informatora, słuchać fragmentu nagranej rozmowy, a także obserwować wizualizację falową plików dźwiękowych. Adnotacjami tworzonymi na wielu warstwach (poziomach) mogą być zdania, wyrazy, opis dowolnej cechy zaobserwowanej w nagraniu. Warstwy te mogą być połączone, np. warstwę uwag i objaśnień wyrazów podporządkowuje się warstwie głównej, niezależnej, w której odnotowywane są wypowiedzi informatora. Konfiguracja i liczba poziomów (niezależnych i podporządkowanych) może być dowolna. To użytkownik przystosowuje warstwy do celu badań. Utworzone anotacje



Ilustracja 1. Transkrypcja wywiadu w oprogramowaniu ELAN. Szablon uwzględniający warstwy informatora, eksploratora, uwagi, objaśnienia oraz wątpliwości

Źródło: ELAN, Instytut Psycholingwistyki im. Maxa Plancka w Holandii.



Ilustracja 2. Inny tryb przeglądania transkrypcji pokazanej na ilustracji 1

Źródło: ELAN, Instytut Psycholingwistyki im. Maxa Plancka w Holandii.

jest na segmenty, które zawierają – na ile to możliwe – spójną semantycznie i syntaktycznie prostą lub złożoną wypowiedź respondenta. Segment warstwy informatora skorelowany jest z innymi warstwami, w których odnotowuje się uwagi (np. o zaistniałej podczas wywiadu sytuacji, śmiechu informatora, wtrąceniach z innych języków), objaśnia wyrazy gwarowe, a także informuje o wątpliwościach związanych z zapisem. Do znakowania leksyki i struktur typowych dla gwary spisanej czy odmienności morfologicznych w kodzie gwarowym stosowany jest stały zestaw symboli (#, //, ^, * i inne).

W trakcie transkrypcji uwzględnia się dwa poziomy anotacji⁷: ogólny oraz gwarowy. Na poziomie ogólnym, powstającym na potrzeby taggera, wszystkie zmiany fonetyczne (np. mazurzenie, samogłoski pochylone, rozkład samogłosek nosowych) sprowadzane są do postaci ogólnej, standaryzowanej. Stosowana pisownia uogólniona, ortograficzna umożliwi przeszukiwanie bazy, a wcześniej automatyczną lematyzację⁸ jednostek oraz znakowanie morfosyntaktyczne za pomocą taggera języka polskiego⁹, wykorzystywanego do tworzenia *Narodowego Korpusu Języka Polskiego*. Dla słownictwa dyferencyjnego lematy i kategorie gramatyczne będą wprowadzane ręcznie (por. Grochola-Szczepanek 2017: 114–115). Anotacja gwarowa także zapisywana jest przy użyciu znaków ortografii ogólnej. Zachowuje się jednak cechy wymowy gwarowej, uwzględnia zmiany morfologiczne (np. inny morfem czy paradygmat fleksyjny). Oba poziomy: ogólny oraz gwarowy, zapisywane są w jednej warstwie. Na odmiennych poziomach znajdują się dopiero w wersji ostatecznej korpusu (Grochola-Szczepanek 2017: 115–117; por. Grochola-Szczepanek 2016: 192).

Transkrypcje tekstów i nagrań uzupełni zamieszczony na stronie bazy słownik wyrazów spiskich ułatwiający korzystanie ze zbioru użytkownikom niezającym dialektu. Objaśnione zostaną wyrazy dyferencyjne, charakterystyczne wyłącznie dla gwary, por. *rajbacka* – „tara do prania”, *biglajz* – „żelazko”, *cuchać* – „trzeć, szorować”, *rozkielacos* – „niekiedy, czasem”, bądź wspólne gwarze i językowi ogólnemu, jednak różniące się znaczeniem, por. *dołek* – „rodzaj małego spichlerza”,

można przeglądać w różnych trybach, a przeszukiwanie dokumentów ułatwia wyszukiwarka (por. ELAN, <https://tla.mpi.nl>, dostęp: 26.02.2018; Sloetjes, Wittenburg 2008, dostęp: 15.05.2018). Oprogramowanie umożliwia nie tylko transkrypcję pliku dźwiękowego, służy także do tagowania lub opisywania plików wideo. Mogłoby zatem być wykorzystane np. do badań nad językiem niewerbalnym. W polskich badaniach gwarowych ELAN nie był do tej pory używany. Za granicą natomiast jest wykorzystywany m.in. w projekcie języka rusińskiego, gwary na terenie Rosji (rejon rzeki Ustii).

⁷ Anotacja (adnotacja, znakowanie) to jawna, lecz niezawarta w formie danej jednostki informacja, która „pozwala na dokonywanie kwerend na podstawie określonych w nich konkretnych cech wyrazów lub tekstów”. Informacją typu językoznawczego może być np. kategoria gramatyczna (Piotrowski, Grabowski 2013: 119).

⁸ Lematyzacja polega na przypisaniu fleksemom (formom wyrazowym) odpowiedniej głównej postaci danego leksemu, czyli jednostki słownikowej.

⁹ Program TaKIPI (Tager Korpusu IPI PAN) ustala opis morfosyntaktyczny słowofrom (form wyrazowych) oraz określa właściwą interpretację wyrazów zależnie od kontekstu ich wystąpienia (Piasecki 2007: 151–152).

fartuch – w stroju spiskim: „zapaska”, *grzać się* – o świni: „odczuwać popę”, *góralski* – „podhalański”.

Możliwości wyszukiwarki planowanego serwisu internetowego

Ostatni etap prac to utworzenie serwisu internetowego, który umożliwi przeszukiwanie bazy nagrań i tekstów przez użytkowników za pomocą różnego typu kwerend prostych lub złożonych. W wyszukiwaniu podstawowym kwerendę możemy przeprowadzić, wpisując token¹⁰ (program szuka wówczas każdej występującej w tekście odpowiedniej formy wyrazowej) lub lemat¹¹ (program znajduje wszystkie formy danego leksemu). Są także pola umożliwiające wpisanie formy gwarowej bądź tagów gramatycznych w formie skrótów¹². Wyszukiwarka pozwala na łączenie zapytań. W polu *Leksem* można również wpisywać początek lub koniec wyrazu (przy użyciu wyrażeń regularnych), dzięki czemu otrzymujemy wszystkie słowa mające dany początek bądź koniec. Kwerenda ta może być połączona z wpisaniem tagu gramatycznego. Jest także możliwość wyszukiwania sekwencji słów lub złożonych ciągów w jednej kwerendzie przez dodanie kolejnego wiersza zawierającego pola wyszukiwania¹³. Istotne jest ułożenie takiego zapytania, by pozwoliło na uzyskanie najlepszych rezultatów wyszukiwania. Szczegółowy opis zawierający metodę tworzenia zapytań w trybie podstawowym i zaawansowanym opublikowany będzie w dostępnej online ostatecznej wersji elektronicznej bazy.

W korpusowym badaniu językowym nie bez znaczenia pozostają frekwencja, waga oraz proporcja badanych zjawisk językowych, uwzględnić należy także dane statystyczne w zależności od zebranego materiału (Hebal-Jeziarska 2013: 26). Przetwarzanie danych „[...] siłą rzeczy nabiera kształtu analizy jakościowości” (Hebal-Jeziarska 2013: 18), co od użytkownika wymaga zastosowania odpowiedniej metody badawczej w pracy z materiałem oraz samym narzędziem (Hebal-Jeziarska 2013: 17). Wybór metodologii i opisu wyekscerpowanego materiału determinuje jednak nie tylko ilość dostępnego materiału, lecz także dziedzina badawcza oraz rodzaj korpusu. Baza spiskich tekstów i nagrań gwarowych może być uznana za korpus specjalistyczny¹⁴, gdyż stanowić będzie „próbę odwzorowania języka używanego przez określoną grupę ludzi w określonych sytuac-

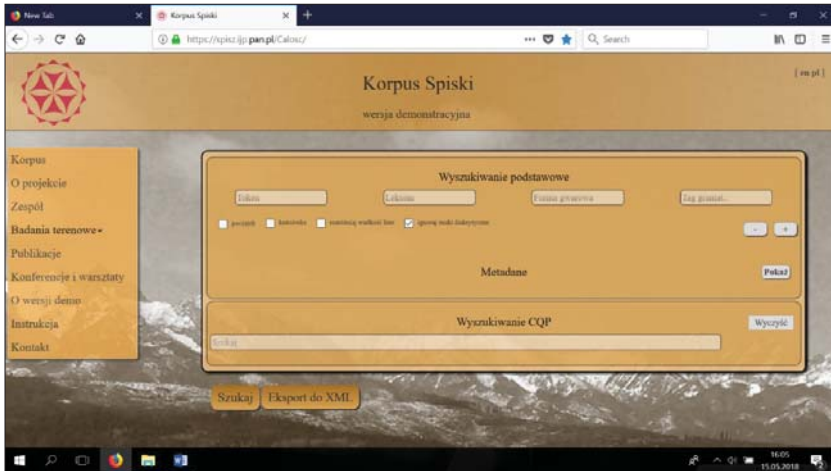
¹⁰ Token – inaczej słowo tekstowe, słowoforma, forma wyrazowa.

¹¹ Lemat (lemma) – główna postać leksemu.

¹² Zbiór znaczników morfosyntaktycznych używany w *Narodowym Korpusie Języka Polskiego*.

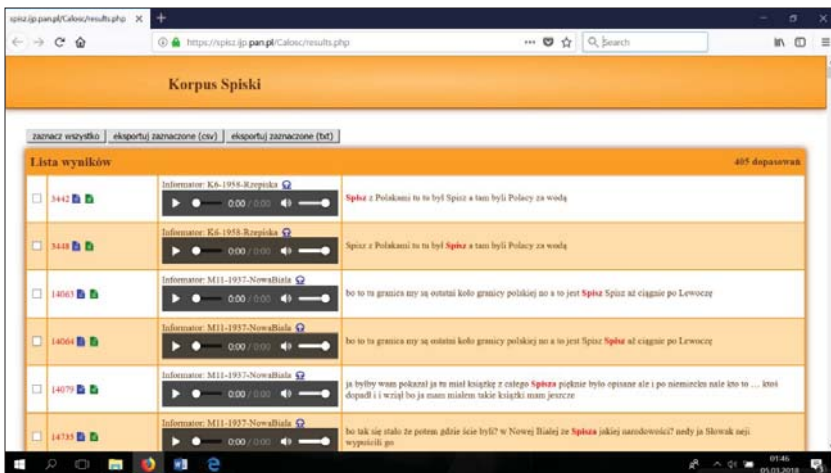
¹³ O sposobach przeszukiwania korpusu zob. m.in. Hebal-Jeziarska 2013; Meger, Woźniak, von Waldenfels 2016.

¹⁴ Termin „język specjalistyczny” jest w pracach językoznawczych różnie definiowany (por. Waliński 2005: 174). W niniejszym artykule przyjęto, że może się także odnosić do gwary danego regionu, jako że jest to jedna z odmian języka narodowego, różniąca się m.in. pod względem lek-



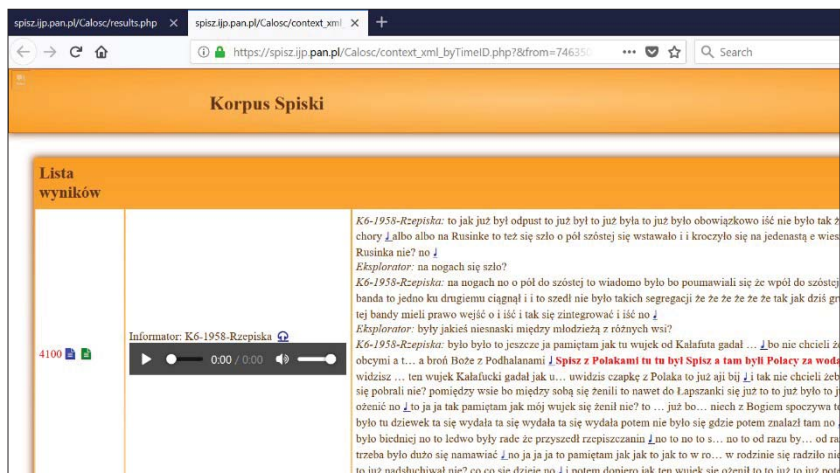
Ilustracja 3. Strona główna wersji demonstracyjnej platformy *Korpus Spiski*.
Wersja demonstracyjna

Źródło: <https://spisz-wew.ijp.pan.pl/Pantera/> (dostęp: 29.06.2018, adres ostatecznej wersji korpusu ulegnie zmianie).



Ilustracja 4. Poświadczenia uzyskane po wpisaniu kwerendy *Spisz* w polu *Leksem*, *Korpus Spiski*.
Wersja demonstracyjna

Źródło: <https://spisz-wew.ijp.pan.pl/Pantera/results.php> (dostęp: 29.06.2018).



Ilustracja 5. Rozszerzenie pierwszego poświadczenia leksemu *Spisz* – kontekst wypowiedzi informatora, *Korpus Spiski*. Wersja demonstracyjna

Źródło: https://spisz-wew.ijp.pan.pl/Pantera/context_xml_byTimeID.php?&from=746350&file=K6-1958-Rzepiska&to=751190 (dostęp: 29.06.2018).

cyjach” (Waliński 2005: 29), w tym wypadku przez mieszkańców Spisza w sytuacji półoficjalnej¹⁵. Korpus specjalistyczny nie musi być tak duży jak korpus ogólny, odwzorowujący język na poziomie powszechnym, z pominięciem żargonów czy dialektów (Waliński 2005: 29; Goźdz-Roszkowski 2005: 180). Jego rozmiar zatem często wyklucza podejście statystyczne. Gdy jednak badacz nie jest nastawiony na zaawansowaną analizę statystyczną, ale szuka jedynie potwierdzenia obranego założenia, nieduża liczba poświadczeń ma wartość pozytywną. Mniejszy materiał można bowiem przeczytać i usystematyzować według przyjętych kryteriów.

sykalnym oraz składniowym. Podobnie klasyfikuje korpus rejestrujący dialektalne odmiany języka Halina Karaś (2015: 89).

¹⁵ Rozmowy z informatorami są oficjalnie nagrywane, respondent wyraża także pisemną zgodę na dokumentację wywiadu. Dąży się jednak do tego, by osoby badane używały języka swobodnego, jak najbardziej zbliżonego do tego stosowanego na co dzień. Pomaga w tym znajomość gwary i kultury Spisza przez eksploratorów. Korzystnie na wyniki badań wpływa także fakt, że wśród badaczy są autochtoni (por. Grochola-Szczepanek 2016: 190).

Przykłady ilustrujące możliwości bazy jako źródła wiedzy historycznej, etnograficznej, folklorystycznej

Wypowiedzi informatorów dotyczą na przykład zmian gospodarczych, społecznych, politycznych, jakie zaszły w ciągu ostatnich kilkunastu lat, a także historii wsi, regionu, osób, które zapisały się w pamięci mieszkańców, por.:

- mężczyzna, ur. 1934, Falsztyn

X¹⁶: *opowie pan o swoim dzieciństwie albo o tutejszym dworze o baronie?*¹⁷.

A: *nedy to było Węgro-Austria tu nie było granicy tak jak i teraz /¹⁸ neji on był Austriak a nie wiem jak dostał ten ten ten ten po... ten tu pole w ogóle na na wsi na Falsztyn / dużo nadal wy... jakies husyci wymordowali kościół był tam na dole y i po prostu wymarnowali¹⁹ / a potem ten Jungenfeld Tivadar ... bo on się nazywał Jungenfeld baron Jungenfeld baron Tivadar ... / a ona była Salamonka²⁰ Węgierka bo to było w Węgrzech i on był Austriak ona była Węgierka / ne to b... różnie było tak że on rządził tu Falsztynem i i i ludzie pracowali na niego / jeszcze i mój ojciec pracował najwięcej na Falsztynie taką pole mieli zelarke²¹ tak zwana takie co odrabiali na cztery krowy to robili coś czterysta osiemdziesiąt dni dwoje na pańszczyznę chodziło / jakby nie poszedł da... do do niego na pańszczyznę to i dostał od niego kulę bił / tak że po prostu było bardzo ciężko tu ludziom życie było było ich czternas... czternastu było tych tych zelarzy.*

- kobieta, ur. 1941, Niedzica-Zamek

X: *co było wcześniej we dworze?*

A: *no to był właśnie ten Szlegier ten rządcą tam mieszkał to był jego dom bo ta hrabina mieszkała w zamku nie? a a on tu / noji jeszcze miał ja... widocznie jeszcze jakichś pomocników ja nie wiem bo ja już tego nie pamiętam to ja już to nie wiem / jacy²² on tu siedział i tu ludzie się zganiłi ku temu dworze / i on rozdzielał ten rządcą rozdzielał ludzi gdzie kto ma iść który ku bydłu który dzieindzie do roboty / no to on to rozdzielał ale ja już tego nie wiem / bo ja się już urodziła we wojnę to to ja już te... a po wojnie już nie było bo pańszczyznę znieśli nie? była reforma i ... / tak że ja już tego nie pamiętam to jacy tyle wiem co tata gadali mój tata.*

- kobieta, ur. 1926, Rzepiska (A); mężczyzna, ur. 1948, Rzepiska (B)

A: *może byście sobie poczytali to to pisze tam kie... i uwierzylibyście kielo ludzi wyniszczyli i kielo zabrali im krów i sierot narobili / czterech bielán pochowali gdzieś tam w Gorcach w jednej dziurze*

¹⁶ X = eksplorator; A, B = informatorzy.

¹⁷ Wszystkie fragmenty wypowiedzi informatorów zapisywane są w postaci standaryzowanej (nieuwzględniającej fonetyki gwarowej). W wersji próbnej korpusu, z którego wyekscerpowane zostały przykłady, możliwa jest wyłącznie anotacja na poziomie ogólnym. W wersji ostatecznej korpusu pojawi się dodatkowo możliwość przeglądania tekstu na poziomie gwarowym, choć wypowiedzi także będą zapisywane w postaci ortograficznej. W cytowanych fragmentach objaśniono tylko niektóre wyrazy gwarowe i nazwy.

¹⁸ Ukośna kreska (/) wyznacza granicę segmentu.

¹⁹ *wymarnować* – „spustoszyć”.

²⁰ *Salamonka* – właśc. Ilona Bethlen Salamon, hrabina, właścicielka zamku w Niedzicy.

²¹ *zelarza* – „pańszczyzna”; *zelarz* – „mężczyzna odrabiający pańszczyznę”.

²² *jacy* – „tylko”.

/ i jednemu nie dało tu z tego Ognia²³ bandy jednemu ne takiemu co chodził z nimi nie? chłopu nie dało sumienie spokoju / i zaszedł do Białej i powiedział pójdźcie klęknał na kolana ja wam pokażę gdzie wasze ludzie leżą i wszyscy czterej w jednej dziurze leżeli i przigrzebiyni²⁴ byli ino²⁵ tak lekko / i po czterech latach ich wozili na cmentarz i robili im pogrzeb w Białej ale się rozlatywali ... tak robili / co za Ogień był co za człowiek był bez sumienia za co tych ludzi bili.

B: neji tak rządził i ... / ale to to źle rządził no bo biednych ludzi bił takich co co co co co co co nie miał z nimi nic nie? może go w życiu ani nie znał i ...

A: niewinnych.

B: co co kto nagadał z całego koleże... jak teraz pana znam pan jest na przykład z Krakowa no tam pan zna trochę Kraków jak mnie pan nagada żem idź do tego idź do tamtego i rób z nim porządek i dał neji daje pieniądze i krowy mu zabierzcie / tu du... dużo tak było po wsiach jak był ktoś zły na mnie no to no to albo ja był zły na waszego tatę no to idź tam do tego neji weź tam zrób porządek z nim i neji robił.

- kobieta, ur. 1950, Frydman

X: dlaczego ludzie przeprawiali się na Słowację?

A: no tam chodzili wtedy na przemysł przeważnie bo tutaj było ciężko a stamtąd przynosili to przynosili papierosy to ... prawdopodobnie i handlowali troszkę dlatego żeby przeżyć nie? / ale tutaj mieli akurat trochę na Spiszu spokoju bo byliśmy pod Słowacją a Słowacy dobrze z Niemcami bo no bo historia / w czasie drugiej wojny światowej Słowacja otworzyła granicę przed Niemcami Niemcy przez Słowację przeszli dalej przez Słowację Polska się broniła a oni się przez Słowację przetargali i przeszli / tak że to historia sama skoro nie Słowacy byli w stosunku lojalni do Niemców no to się okazuje że tutaj myśmy na Spiszu jaki taki spokój mieli nie? / no ale to już samo przechodzenie przez Dunajec i przez granicę po to żeby przemysł ... na przemysł iść to to była taka kontrabanda trochę nie? / no to tutaj już nie za bardzo patrzyli ci finance²⁶ na ... bo to byli finance na bardzo ... nie za bardzo patrzyli na tych którzy to robili którzy to wykonywali nie?

Wywiady z mieszkańcami Spisza to również rozmowy na temat tradycji: wierzeń, zwyczajów, obrzędów, zabobonów. Wypowiedzi respondentów pokazują różnice w zwyczajach (ta sama tradycja może przybierać różne formy, mieć nieco inny schemat, np. „ogrywanie majów”²⁷) oraz nakreślają obyczajowość mieszkańców poszczególnych wsi spiskich czy sąsiedniego regionu – Podhala. Stanowią także źródło wiedzy o kulturze materialnej – narzędziach, strojach, budownictwie itp. Zdarza się, że respondent przytacza tzw. „gadki” – podania ludowe, snuje opowieści, wspomina czas wojenny, młodości swojej lub przodków, a nawet śpiewa (usłyszeć można np. przyspiewki weselne). Por.:

²³ Ogień – pseudonim Józefa Kurasia, żołnierza Wojska Polskiego i Armii Krajowej.

²⁴ przigrzebiyni – „przysypani”.

²⁵ ino – „tylko”.

²⁶ financ – „żołnierz straży granicznej; celnik”.

²⁷ „Ogrywanie majów” to zwyczaj chodzenia po wsi orszaku kawalerów, grania i tańczenia przed domami pańien w Zielone Świątki.

- kobieta, ur. 1941, Jurgów

X: ludzie wierzyli w przesady?

A: wierzyli w gusła i też było takich dużo ludzi że wierzyli w gusła a i że umieli porabiać²⁸ dawno ludzie jeden drugiemu / na przykład że ktoś był na ... jeden na drugiego zły to umieli porabiać że ... na zwierzętach najbardziej umieli po... porabiać / jeden w Rzepiskach to mu owce zdechły coś dziesięć czy piętnaście / bo mu tak porobił jeden coś był na na tego na tego człowieka był coś zły niedobry nie? to mu dziesięć owiec zdechło tak umieli ludzie porabiać / no a z ... tam poza Górę tam poza Czarną Górę zaś ludzie umieli ... dużo było takich ludzi co umieli porabiać ludziom ludziom umieli porabiać co głupieli / psychicznie nie? byli potem później chorzy byli głupieli no to było ... też płone²⁹ były te ludzie nie?

- mężczyzna, ur. 1954, Dursztyn

A: to jest zwyczaj taki ogro... ogrywanie majów to jest na Zielone Świątki / zawsze w drugi dzień świąt strażacy z ochotniczej straży wychodzili i był młody pan taki albo przebrany albo tego starosta tego było co gadał takie przemówienie / i szli koło każdego domu i y to jeszcze y bie... y wbijali takie maje tam gdzie była dziewczyna czy dwie czy czy dorodna jakaś tego to ba... bardzo piękny maj dawali / to taki było d... drzewko albo smreczek był albo olszyna liściaste albo o to no to takie było / neji to d... w przeddzień tych Zielonych Świątek jak mieli ogrywać to wbijali maje / a potem koło tych majów przychodzili i ogrywali te maje i y muzyka była na ... y harmonia neji co tam gęśle były wtedy mod... w modzie i i neji harmonia.

- kobieta, ur. 1962, Dursztyn

A: dawno też na Wigilię to były zwyczaj że wyszły dziewczki na pole jak już pojedli wigilię i hujkała³⁰ choćtoro / zahujkała no to gdzie się to echo porozl... rozniosło ona tam niby słyszała no to wiedziała że tam poszedł jej głos i tam ona pójdzie za niewiastę / albo wychodziły i stawiantki³¹ jak ktoś miał ogrodzenie z tego obłapiwały stawiantki że jak na szerokość kielo rękami obejmie to za tyle lat się wyda e to miały takie swoje / do stajni szli słuchali bo krowy tam barany gadają ludzkim głosem ale tam nikt nie słyszał tego nale tradycja się utrzymuje bo gadają / choć ta stryjno moja właśnie co gadam ta ta Bizub ona gada że we Frydmanie słyszał ktoś tam że wlaźł do stajni i te krowy gadały że będzie umarły a drugi stał na polu i on to słyszał i umarł / tak że że było fajnie naprawdę było fajnie dawno takie zwyczaj takie... [...]

X: zaśpiewa pani przyspiewkę?

A: (informatorka śpiewa) pamiętajże Hanuś żebyś ją ... pamiętajże Józus żebyś ją szanował żebyś ją szanował bo dość chłopców było a tyś się sam wmawiał bo dość chłopców było a tyś się sam wmawiał / pamiętajże Maryś byś go nie otrula byś go nie otrula by cię nie nadeszła cygańska natura by cię nie nadeszła cygańska natura / potem na rodziców żeby no tak im dogryźć to stara Pieknoskula w sieni groch cudziyla³² w sieni groch cudziyla o dobrego zięcia Boga uprosiła o dobrego zięcia Boga uprosiła / stara Pieknoskula noże pochowała noże pochowała żeby jej niewiasta chleba nie krajała.

²⁸ porabiać – „czarować, rzucać uroki, klątwy”.

²⁹ płony – „kiepski, marny, lichy”.

³⁰ hujkać – „pokrzykiwać, wydając przeciągły głos”.

³¹ stawiantka – „sztachetka”.

³² cudzić – „przesiewać”.

Zebrane teksty gwarowe ujawniają ponadto tryb życia i kulturę wspólnoty wiejskiej w jej naturalnym środowisku. Informatorzy wielokrotnie dzielą się refleksjami na temat otaczającej ich rzeczywistości oraz zachowań i postaw społeczności, której są częścią. Codziennosc odzwierciedlona w wypowiedziach respondentów może, podobnie jak obserwacja, posłużyć do badań nad zachodzącymi współcześnie zmianami społecznymi. Por.:

- kobieta, ur. 1972, Trybsz

X: *jak wyglądały dawniej relacje rodzice–dzieci? czy rodzice byli opiekuńczy w stosunku do dzieci?*
 A: *dawno rodzice właśnie ja powiem ... teraz teraz te dzieci są bardziej wychowywane no nie? dawno dzieci musiały iść w pole / jak były małe ... dziecko to rodzic musiał brać dziecko i iść czy płakało mrajało to trzeba było iść w pole a dziś jak ma młoda ... jak tam teściowie jeszcze gazdują młode jak ma to ona siedzi w domu z dzieckiem no nie? a a rodzice idą / tak że dziś są dzieci tak wycackane zależy jak jak ktoś z domu wyniesie / bo są dzieci co szanują dziadków a są dzieci co ... wie pani co co co za nic ich nie mają no nie? / no tak jak u nas to my tutaj nauczeni bo moi rodzice no nie? jeszcze do niedawna no to my dziadki dziadki a są co tam w ogóle o to wiesz jak jak to się dziś gada no nie? no / dawno wie pani co powiem tylko tyle że dawno ludzie się ... jak szedł ktoś z pola wiedział że ma na przykład dużo składać kopki czy ma dużo kosić to każdy się obracał i pomagał każdemu a dziś? / jakbyś był w polu może ci grzmieć tak my rękami to objedzie coś cię Boże broń żeby ci kto pomógł to nie ma takiej opcji / bo dziś jest taka zawiść między ludźmi że ... żeby ino ten miał lepiej żeby ciebie ni... nie było lepiej jak jak komuś tak że naprawdę / tu w Trybszu jest okropnie bo ja to tak słucham co takie obce jeżdżą do nas to oni gadają że oni mogą iść do sąsiada wziąć traktor czy auto a my? jakby ci traktor to możesz ... tam ci może zgnić siano ale co ty nie ma że pożyczysz od kogo.*

- mężczyzna, 1962, Niedzica

A: *no ale ja myślę tak że no w tej chwili na świecie jest bardzo dobrze w tej chwili na świecie się ludziom żyje bardzo dobrze / nie ma się co oszukiwać no żyje się żyje się w tej chwili bardzo dobrze jest wszystkiego jest w sklepach no tak jak jak jeszcze jak mówię jak jest zdrowie i i pieniądz to jest wszystko nie? to jest wszystko / no dziś ludzie młodzi dużo tak samo dawniej wyjeżdżali za granicę nie? bo tak samo szli do Pesztu na pracę i ci którzy do niej szli a w tej chwili no dawniej szli pieszo tam koniem gdzie wyjeżdżali tego ale a w tej chwili jest tak że siądzie w samolot autobus leci do Anglii / noji noji są inne pieniążki nie? noji ci ludzie młodzi uciekają za granicę po prostu a u nas w Polsce no no jest jak jest i żeby tych ludzi młodych zatrzymać no coś im trzeba obiecać no ale ja nie wiem co im obiecać żeby żeby zostali żeby byli w Polsce nie?*

Podsumowanie

„Korpus tekstów i nagrań z polskiego Spisza będzie dokumentacją języka mieszkańców wsi w jego naturalnym, wiejskim środowisku na początku XXI wieku” (Grochola-Szczepanek 2017: 118). Baza transkrypcji nagrań dialektalnych ze względu na szeroki zakres tematyczny wywiadów, dostępność interfejsu korpusu oraz optymalną liczbę tekstów może być jednak podstawą bądź uzupełnieniem

badań (np. ankiet, obserwacji, analiz źródeł prymarnie pisanych) dla specjalistów wielu dyscyplin naukowych, nie tylko dialektologów, etnolingwistów czy też socjolingwistów. Współzależność języka i kultury pozwala bowiem na korzystanie z tego samego materiału badawczego, choć z zastosowaniem odrębnej metodologii.

Bibliografia

Biały Z.

1987 *Polski Spisz. Historyczne uwarunkowania kultury ludowej tego regionu i sposoby jej badania po II wojnie światowej*, w: Z. Biały (red.), *Polski Spisz. Jedność kultury ludowej i jej historyczne uwarunkowania. Środowisko naturalne – warunki gospodarowania. Antropologia. Gwary*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace Etnograficzne”, nr 22, s. 7–57.

Gajda S.

2006 *Komunikacyjny model zróżnicowania współczesnej polszczyzny*, w: J. Sierociuk (red.), *Gwary dziś. 3: Wewnętrzne zróżnicowanie języka wsi*, Poznań, s. 7–17.

Goźdz-Roszkowski S.

2005 *Korpusy a języki specjalistyczne*, w: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Łódź, s. 174–200.

Grochola-Szczepanek H.

2006a, *Badania fokusowe mowy mieszkańców wsi*, „Socjolingwistyka”, nr 20, s. 19–35.

2006b, *Wpływ przemian społecznych i kulturowych na mowę mieszkańców wsi (na przykładzie wsi Rzepiska)*, w: J. Sierociuk (red.), *Gwary dziś. 3: Wewnętrzne zróżnicowanie języka wsi*, Poznań, s. 167–174.

2012 *Język mieszkańców Spisza. Płeć jako czynnik różnicujący*, Kraków, [http://rcin.org.pl/Content/49559/J%3C3%84%E2%84%A2zyk%20mieszka%C4%B9%E2%80%9Ec%C4%82%C5%82w%20Spisza%20\(18%20IX%202014\).pdf](http://rcin.org.pl/Content/49559/J%3C3%84%E2%84%A2zyk%20mieszka%C4%B9%E2%80%9Ec%C4%82%C5%82w%20Spisza%20(18%20IX%202014).pdf) (dostęp: 16.05.2018).

2013 *Badania języka mieszkańców wsi w kontekście przemian społecznych*, „Socjolingwistyka”, nr 27, s. 43–53.

2016 *Korpusowe badania języka mieszkańców Spisza w Polsce – cele i zadania*, „Jezi-koslovni Zapiski. Zbornik Inštituta za Slovenski Jezik Frana Ramovša ZRC SAZU”, nr 22(2), s. 185–196.

2017 *Nowe badania języka mieszkańców wsi regionu polskiego Spisza*, w: B. Osowski, J. Kobus, P. Michalska-Górecka, A. Piotrowska-Wojaczyk (red.), „Poznańskie Studia Polonistyczne Serii Językoznawczej”, nr 47, Poznań, s. 103–119.

Hebal-Jeziarska M.

2013 *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Opole, s. 17–30.

Karaś H.

2014 *O planowanym Korpusie Gwar Polskich*, w: M. Rak i K. Sikora (red.), *Badania dialektologiczne. Stan, perspektywy, metodologia. Materiały konferencji naukowej „Gwara i tekst”*, Kraków, 27–28 września 2013 r., Kraków, s. 33–43.

2015 *Zadania dialektologii w zakresie ochrony ginących gwar*, w: J. Sierociuk (red.), *Gwary dziś. 7: Rocznik poświęcony dialektologii słowiańskiej*, Poznań, s. 83–95.

- Karaś H. (red.),
2010 *Dialekty i gwary polskie. Kompendium internetowe*, Zakład Historii Języka Polskiego i Dialektologii UW, Towarzystwo Kultury Języka, <http://www.dialektologia.uw.edu.pl/index.php?l1=start> (dostęp: 15.05.2018).
- Kąś J.
2001 *Metodologia badań leksyki gwarowej w kontekście współczesnych przemian kulturowych i społecznych*, w: J. Sierociuk (red.), *Gwary dziś. 1: Metodologia badań*, Poznań, s. 191–200.
- Krawczyk-Wieczorek A.
2012 *Automatyczna lematyzacja tekstu w zapisie fonetycznym. Korpus polskiej gwary południowokresowej*, „Język Polski”, nr 1, s. 11–19.
- Kurek H.
1995 *Przemiany językowe wsi regionu krośnieńskiego (studium socjolingwistyczne)*, Kraków.
2005 *Gwary dziś*, w: M. Czermińska, S. Gajda, K. Kłosiński, A. Legeżyńska, A.Z. Makowiecki, R. Nycz (red.), *Polonistyka w przebudowie. Literaturoznawstwo – wiedza o języku – wiedza o kulturze – edukacja. Zjazd Polonistów. Kraków, 22–25 września 2004*, t. 1, Kraków, s. 589–595.
2006 *Zróżnicowanie polszczyzny wiejskiej a czynniki pozajęzykowe*, w: J. Sierociuk (red.), *Gwary dziś. 3: Wewnętrzne zróżnicowanie języka wsi*, Poznań, s. 53–57.
- Lewandowska-Tomaszczyk B.
2005 *Powstanie i rozwój językoznawstwa korpusowego*, w: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Łódź, s. 9–26.
- Meger A., Woźniak M., Waldenfels von R.
2016 *Jak stworzyć korpus równoległy „dla wszystkich”? O pracy nad Polsko-Niemieckim i Niemiecko-Polskim Korpusem Równoległym*, w: E. Gruszczyńska i A. Leńko-Szymańska (red.), *Polskojęzyczne korpusy równoległe. Polish language Parallel Corpora*, Warszawa, s. 97–118.
- Pelcowa H.
2002 *Dialektologia wobec wyzwań XXI wieku*, w: S. Gala (red.), *Dialektologia jako dziedzina językoznawstwa i przedmiot dydaktyki. Księga jubileuszowa dedykowana Profesorowi Karolowi Dejniu*, Łódź, s. 383–392.
2012 *Dialektologia – stan i perspektywy badań*, w: H. Pelcowa i M. Wojtak, *Słowa. Style. Metody*, Lublin, s. 205–216.
- Piasecki M.
2007 *Polish Tagger TaKIPI: Rule Based Construction and Optimisation*, „Task Quarterly”, nr 11, 151–167.
- Piotrowski T., Grabowski Ł.
2013 *Autorski glosariusz wybranych terminów. Hasło: adnotowanie*, w: W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Opole, s. 119–122.
- Radomski A., Bomba R.
2013 *Zwrot cyfrowy w humanistyce*, w: A. Radomski, R. Bomba (red.), *Zwrot cyfrowy w humanistyce. Internet/Nowe Media/Kultura 2.0*, Lublin, s. 7–9, http://e-naukowiec.eu/wp-content/uploads/2013/05/Zwrot_cyfrowy_w_humanistyce.pdf (dostęp: 1.03.2018).
- Sloetjes H., Wittenburg P.
2008 *Annotation by Category – ELAN and ISO DCR*, referat przedstawiony na konferencji *Proceedings of the 6th International Conference on Language Resources and*

Evaluation, Marrakesz, http://pubman.mpdl.mpg.de/pubman/item/escidoc:60774:3/component/escidoc:60775/Sloetjes_2008_annotation.pdf (dostęp: 14.05.2018).

Waliński J.

2005 *Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej*, w: B. Lewandowska-Tomaszczyk (red.), *Podstawy językoznawstwa korpusowego*, Łódź, s. 27–41.

źródła internetowe

Dialektologiczna Fonoteka Wydziału Polonistyki Uniwersytetu Jagiellońskiego, 2014–2016, <http://www.fonoteka.polonistyka.uj.edu.pl> (dostęp: 15.05.2018). *ELAN*, <https://tla.mpi.nl> (dostęp: 26.02.2018).

Korpus Spiski. Wersja demonstracyjna, 2015–2018, <https://spisz.ijp.pan.pl/Calosc> (dostęp: 15.05.2018).

Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 13 grudnia 2013 r. w sprawie ustalenia, zmiany i zniesienia urzędowych nazw niektórych miejscowości oraz ustalenia nazwy obiektu fizjograficznego, Dziennik Ustaw Rzeczypospolitej Polskiej, Warszawa 2013, poz. 1629, <http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20130001629/O/D20131629.pdf> (dostęp: 26.02.2018).