

Danuta Roszko

Eksperymentalny korpus litewskiej gwary puńskiej w Polsce (narzędzie do badań nad gwarami polskimi i litewskimi na Suwalszczyźnie)

0. Wstęp

Lingwistyka korpusowa to dziedzina nauki, której rozwój z roku na rok nabiera tempa. Początkowo zainteresowania badaczy były skoncentrowane na opracowywaniu korpusów jednojęzycznych (tzw. narodowych), później – dwu- i wielojęzycznych korpusów równoległych. W ostatnim okresie uwaga lingwistów komputerowych zostaje skierowana na korpusy gwar i dialektów, por.: (a) Nordic Dialect Corpus (<http://www.tekstlab.uio.no/nota/scandiasyn/index.html>), (b) Helsinki Dialect Corpus (http://blogs.helsinki.fi/hes-eng/files/2011/03/HES_Vol2_Peitsara_Vasko.pdf), (c) FRED – Freiburg Corpus of English Dialects (<http://www.helsinki.fi/varieng/CoRD/corpora/FRED/index.html>), (d) Estonian Dialect Corpus (<http://www.murre.ut.ee/estonian-dialect-corpus/>), (e) YADAC – Dialectal Arabic Corpus (http://www.lrec-conf.org/proceedings/lrec2012/pdf/663_Paper.pdf), (f) Catalan Corpus Oral Dialectal (<http://www.uv.es/foncat/cat/Treballs/10.Clua-Lloret.pdf>), (f) Russian National Corpus (Dialectal corpus) (<http://www.ruscorpora.ru/en/corpora-structure.html>) i in.

Twórcy korpusów gwarowych stoją przed trudnym zadaniem. Najważniejszym problemem jest ograniczony dostęp do zasobów gwarowych. Zazwyczaj sytuacja wymaga zebrania materiału w terenie. Jest to zadanie odpowiedzialne i długotrwałe. Niezmiernie ważny jest właściwy dobór informatorów według kryteriów pokolenia, płci, wykształcenia, zawodu i in. Nie mniej

istotne jest odpowiednie zrównoważenie tematyczne zapisywanych rozmów, co bezpośrednio wpływa na bogactwo form leksykalnych, morfologicznych i konstrukcji składniowych.

Po skompletowaniu nagrań gwarowych następuje kolejny etap prac, jakim jest przetworzenie zebranych materiałów do postaci cyfrowej (np. do plików tekstowych TXT w standardzie UTF8). Ten etap wymaga ustalenia formalnych kryteriów zapisu (fonetyczny czy uproszczony, wzorowany na języku ogólnym normowanym) oraz systemu kodowania. Przyjęte założenia rzutują na kolejny etap prac nad korpusem. Jest nim lematyzacja i anotacja zasobów oraz organizacja metadanych (zawierających notki o informatorach, miejscu, dacie nagrania i in.). Do przeprowadzenia lematyzacji i anotacji tekstu gwarowego konieczne może okazać się stworzenie od podstaw programu automatyzującego te procesy lub trenowanie istniejących taggerów (jeśli taką opcję zapewniają). Wynik automatycznie przeprowadzonej lematyzacji i anotacji wymaga pracochłonnej ręcznej korekty. Ostatnim etapem jest załadowanie plików do programu zarządzającego korpusem.

1. Eksperymentalny korpus gwary puńskiej w Polsce (EKorpGP)

W roku 2010 przystąpiono do budowy korpusu¹. Wykorzystano materiały pochodzące z nagrań zarejestrowanych (przez piszącą te słowa) podczas wielu ekspedycji terenowych, poczynając od 1986 roku. Część włączonych do korpusu tekstów pochodzi z rozmów, o których nagrywaniu rozmówcy zostali uprzedzeni. Większa część została zarejestrowana spontanicznie w miejscach publicznych. Należy podkreślić, że korpus jest nieustannie rozbudowywany. Stopniowo są dodawane kolejne przetworzone do plików tekstowych nagrania dźwiękowe. Szerzej na temat samego korpusu, jego powstania, struktury, zastosowanego oprogramowania, typu anotacji, por. Roszko 2013.

Z końcem 2012 roku objętość korpusu osiągnęła blisko 260 000 tokenów². Liczba lematów przekroczyła wartość szesnastu tysięcy³. Łącznie w korpusie

¹ Prace nad EKorpGP prowadzone są przez autorkę tego artykułu. W kwestiach technicznych autorka korzysta z pomocy R. Roszki, z którym wspólnie opracowuje Eksperymentalny korpus polsko-litewski.

² Przykłady form uwzględnionych w obliczeniach: *namas* 'dom' (forma liczby pojedynczej w mianowniku), *namai* 'jw.' (forma liczby mnogiej w mianowniku), *namų* 'jw.' (forma liczby mnogiej w dopełniaczu), *aš* 'ja' (forma liczby pojedynczej w mianowniku), *manim* 'jw.' (forma liczby pojedynczej w narzędniku) itd.

³ Przykłady lematów uwzględnionych w obliczeniach: *namas* 'dom', *aš* 'ja', *vienas* 'jeden' itd.

zanotowano 68 411 unikatowych form. Podane wartości odnoszą się do głównych zasobów korpusu, umownie nazwanych podkorpusem A (por. niżej p. 2.1).

2. Zasoby EKorpGP

Materiały dotychczas włączone do korpusu zostały podzielone na trzy formalne grupy zwane podkorpusami A, B i C.

2.1. Podkorpus A – to jednojęzyczny korpus zawierający wypowiedzi Litwinów puńskich gwarą, por. Tabela 1. Dane statystyczne o podkorpusie A zostały podane wyżej w p. 1.

Tabela 1. Fragment rozmowy włączony do zasobów podkorpusu A wraz z tłumaczeniem na język polski (Materiał pochodzi z 2011 roku).

Lp.	Podkorpus A	Tłumaczenie na język polski
1.	[K15] [...] nu tai Vaitakiemio uždarė.	[...] no to w Wojtokiemiach zamknęli [szkołę]
2.	[K45] Ir Navinykuose, kap VP pradėjo dirbti. Taigi jisai dešimt metų gal virš važinėję in Navinykus, da vis jis turėdavo pusę etato, Punkski dirbo. Nu tai jis už mani dzviem metais jaunesnis. Tai devyniasde- ... apie šimtas vaikų buvo Navinykų mokykloj, o dabar gal trisdešimt likę.	I w Nowinikach, jak VP zaczął pracować. To on może więcej niż dziesięć lat jeździł do Nowinik, jeszcze miał pół etatu, w Puńsku pracował. No, on ode mnie jest o dwa lata starszy. To dziewięćdzie- ... około stu dzieci było w szkole w Nowinikach, a teraz może trzydzieści zostało.
3.	[K15] Ir daugiausia iš visų kaimo mokyklų yra Navinykuose dabar. Pristavonyse tai gal penki, šeši.	I najczęściej [dzieci] ze wszystkich szkół wiejskich jest teraz w Nowinikach. W Przystawańcach to jest może pięciu, sześciu [uczniów] .
4.	[K45] Nežinau, devyni tį buvo...	Nie wiem, chyba dziewięciu było.
5.	[K15] Ale mažai labai. Koki vienas antroji klasėj, du tračioji klasėj ir tep va.	Ale bardzo mało. Jakoś tam jeden w drugiej klasie, dwóch – w trzeciej klasie i tak to.

Legenda:

[K15] – nastolatka, lat 15, gimnazjalistka, zam. w okolicach Puńska

[K45] – kobieta, lat 45, wykształcenie wyższe, zam. w okolicach Puńska

W oparciu o przytoczony w tabeli 1. fragment wypowiedzi przedstawicielki dwóch pokoleń (młodszej i średniej) można stwierdzić, że liczba cech

gwarowych jest zdecydowanie wyższa w wypowiedzi kobiety reprezentującej pokolenie średnie. Wypowiedź gimnazjalistki wykazuje znacznie mniejsze nacechowanie elementami gwarowymi.

Materiał zawarty w podkorpusie A został opracowany zgodnie z zasadami obowiązującymi w lingwistyce korpusowej. Wszystkim formom zostały przypisane lematy oraz wykaz charakterystycznych cech morfologicznych i syntaktycznych, por. Tabela 2. i 3.

Tabela 2. Lematy i wybór cech morfosyntaktycznych na przykładzie zdania *Ir daugiausia iš visų kaimo mokyklų yra Navinykuose dabar.* (por. Tabela 1, wiersz 3).

Lp.	Token	Lemat	Opis cech morfosyntaktycznych
1.	Ir	ir	jng.
2.	daugiausia	daugiausia	prv., teig., nelygin. l.
3.	iš	iš	prl.
4.	visų	visas	įv., mot. g., dgs., K.
5.	kaimo	kaimas	dkt., vyr. g., vns., K.
6.	mokyklų	mokykla	dkt., mot. g., dgs., K.
7.	yra	būc	vksm., teig., nesngr., tiesiog. n., es. l., dgs., 3 asm.
8.	Navinykuose	Navinykai	dkt., vyr. g., dgs., Vt.
9.	dabar	dabar	prv., teig., nelygin. l.

Legenda:

Oznaczenie	Treść	Oznaczenie	Treść
3 asm.	osoba 3.	nesngr.	forma niezwrótana
dgs.	liczba mnoga	prl.	przyimek
dkt.	rzeczownik	prv.	przysłówek
es. l.	czas teraźniejszy	teig.	forma pozytywna
įv.	zaimek	tiesiog. n.	tryb oznajmujący
jng.	spójnik	vyr. g.	rodzaj męski
K.	dopełniacz	vksm.	czasownik
mot. g.	rodzaj żeński	vns.	liczba pojedyncza
nelygin. l.	stopień równy	Vt.	miejsownik

Tabela 3. Przykład anotacji morfosyntaktycznej zdania *Ir daugiausia iš visų kaimo mokyklų yra Navinykuose dabar*⁴.

```

<word="Ir" lemma="ir" type="jng."/>
<space/>
<word="daugiausia" lemma="daugiausia" type="prv., teig., nelygin. l."/>
<space/>
<word="iš" lemma="iš" type="prl."/>
<space/>
<word="visų" lemma="visas" type="iv., mot. g., dgs., K."/>
<space/>
<word="kaimo" lemma="kaimas" type="dkt., vyr. g., vns., K."/>
<space/>
<word="mokyklų" lemma="mokykla" type="dkt., mot. g., dgs., K."/>
<space/>
<word="yra" lemma="būc" type="vksm., teig., nesngr., tiesiog. n., es. l., dgs., 3
asm."/>
<space/>
<word="Navinykuose" lemma="Navinykai" type="dkt., vyr. g., dgs., Vt."/>
<space/>
<word="dabar" lemma="dabar" type="prv., teig., nelygin. l."/>
<sep="."/>

```

W tabeli 2. w sposób poglądowy przedstawiono przykłady lematów oraz wykaz cech morfosyntaktycznych dla wszystkich form wybranego zdania z podkorpusu A. W tabeli 3. natomiast znajduje się dokładna kopia fragmentu pliku wynikowego, który, pozostając niewidocznym dla użytkownika, wpływa na poprawność uzyskiwanych odpowiedzi dla funkcji poszukiwania określonych danych w korpusie. Na przykład, po zadaniu pytania o formy rodzaju żeńskiego, z podawanego jako przykład w tabeli 2. i 3. zdania zostaną wyświetlone formy *visų* i *mokyklų* wraz z kontekstem.

2.2. Podkorpus B – to jednojęzyczny korpus zawierający wypowiedzi Litwinów puńskich po polsku. Na początkowym etapie powstawania EKorpGP nie brano pod uwagę utworzenia korpusu zawierającego wypowiedzi Litwinów po polsku. Z czasem zdano sobie sprawę, że jest to również materiał cenny, który może zostać wykorzystany nie tylko do badań nad polszczyzną Litwinów

⁴ Wykorzystano tagger *Morphological annotator* (<http://tekstynas.vdu.lt/page.xhtml?id=morphological-annotator>).

puńskich, lecz również do badań porównawczych polszczyzny „w wariacie litewskim” w zestawieniu z okolicznymi gwarami polskimi Suwalszczyzny. W tabeli 4. znajduje się przykład wypowiedzi Litwina, kierowanej do grupy turystów z centralnej Polski.

Tabela 4. Fragmenty zaczerpnięte z podkorpusu B.

[M60] Tutej z naszej strony to nie było żadnych patroli, a tu z Litwy strony, nie, tutaj były był patrol, tu wszystko było przywiezione tam te azjaci. Oni byli tak nastawieni ... Tak byli nastawiony, że za granico to sami vrogovie tutej mieszkajo. Zaras tutaj taki był nauczycielem, dyrektorem i sekretarzem gminnym był, jak kiedyś. To były siedemdziesiąte jakieś drugi rok, jakiś tak o mniej viencej. [...]

Legenda:

M60: Mężczyzna, lat 60, rolnik, wykształcenie podstawowe, zam. w Puńsku (nagranie z 2010 r.)

[M76] Może mnie łapcie tak z dalej, to ładniejszy bendy. Z bliska to już cała starość widać.

Legenda:

M76: Mężczyzna, lat 70, rolnik, wykształcenie 4 klasy, zam. w okolicach Puńska (nagranie z 1992 r.)

Na początkowym etapie prac nad korpusem należało zdecydować, czy w zapisie materiału będzie stosowana transkrypcja fonetyczna, czy też zapis uproszczony. Ze względu na fakt, że aspekty fonetyczne gwary puńskiej zostały już opracowane⁵, w EKorpGP przyjęto zapis uproszczony, wzorowany na pisowni języka normowanego i zarazem znany z literatury przedmiotu, por. między innymi Smoczyński 1984a-b, 1986a-b; Krištopaitė 1998, 1999. W związku z tym analogiczny wariant zapisu przyjęto również dla podkorpusu B. Wprowadzone do normatywnej pisowni polskiej zmiany mają oddawać specyfikę wymowy Litwinów puńskich. Na przykład, stosuje się zapis *v* zamiast polskiego *w* nie tylko dlatego, by zasugerować, że jest to spółgłoska

⁵ Dotychczas gwara puńska nie była badana pod kątem zdaniowych kategorii semantycznych, funkcji składniowych. Ze względu na fakt, że te zagadnienia szczególnie interesują autorkę, nie uznano transkrypcji fonetycznej za nieodzowny atrybut tworzonego korpusu. Należy nadmienić, że włączenie zapisu fonetycznego do korpusu w znaczący sposób wpłynęłoby na dalsze etapy konstruowania samego korpusu. Nie ma bowiem ogólnie dostępnych narzędzi programowych, które byłyby w stanie analizować materiał zapisany fonetycznie. Z kolei korpus gwarowy bazujący na zapisie uproszczonym jest dostępny dla szerszego grona badaczy, czytelników.

wargowo-zębowa, lecz przede wszystkim, że nie podlega ubezdźwięcznieniu. Polskojęzyczny czytelnik widząc *w*, na przykład w wyrażeniu z *Litwy*, automatycznie czyta tę formę jak z *Litfy*. Jednak gdy w miejscu spodziewanego *w* widzi *v*, odruchowo sięgnie do zasad zastosowanej w korpusie pisowni, by czegoś więcej dowiedzieć się na ten temat.

Zasoby podkorpusu B uznaje się za przydatne do badań językoznawczych (wspomniane wcześniej badania różnych wariantów polszczyzny na Suwalszczyźnie), socjologicznych, kulturoznawczych, historycznych, etnologicznych. Z pewnością zasoby podkorpusu B zainteresują badaczy językowego obrazu świata, także polityków zajmujących się sprawami mniejszości narodowych w Polsce.

2.3. Podkorpus C – to dwujęzyczny polsko-litewski korpus równoległy, zawierający teksty pierwotne litewskie/gwarowe i wtórne polskie. Do podkorpusu C zostały włączone tylko te teksty, które są wypowiedziami Litwinów puńskich w gwarze (lub po litewsku) oraz przekłady tych tekstów na język polski, których tłumaczami są również Litwini – mieszkańcy Puńska i okolic. Część zasobów podkorpusu C pochodzi z internetu, część zaś – została zarejestrowana w trakcie spotkań, podczas których tłumaczono wypowiedzi litewskojęzyczne na język polski. W tabeli 5. przedstawiono fragment zasobów podkorpusu C.

Tabela 5. Przykład tekstów włączonych do zasobów podkorpusu C.

Lp.	Wersja litewskojęzyczna	Wersja polskojęzyczna
1.	[...] Jotvingius XIII a. pradžioje nukariavo kryžiuočiai. Tačiau šio krašto nepajėgta intensyviai apgyvendinti, todėl iki XV a. čia ošė miškai. Tik XV a. pradžioje pradėta naujai šiame krašte kurtis. Naujieji krašto šeimininkai buvo lietuviai, kilę nuo Merkinės ir Punios. Jie čia tyvuliuojantį ežerą ir pavadino Punia, nuo kurio ir gyvenvietė gavo vardą. [...]	[...] Jaćwingów na początku XIII w. podbili krzyżacy*, ale ich ziem nie byli w stanie zaludnić, dlatego też do XV w. porastała je puszcza. Dopiero na początku XV w. pojawiają się tu nowi osadnicy. Nowymi gospodarzami ziemi puńskiej stali się Litwini wywodzący się z nad Merecza i Puni. To oni miejscowe jezioro nazwali Punia, od którego później nazwę przejęła także osada. [...]
2.	http://punkskas.pl/pkv2-pl.htm	http://punkskas.pl/?page_id=18

* Pisownia zgodna z oryginałem.

Przedstawiony w tabeli 5. urywek pochodzi z internetu. Z pewnością są to teksty redagowane, jednak potknięcia językowe są możliwe w obu wariantach.

tach językowych. W wierszu 2. (por. tabela 5.) zostały podane adresy źródeł, z których zaczerpnięto teksty. W zasobach korpusowych ta informacja zostaje włączona do plików z metadanymi.

Poniżej w tabeli 6. został przytoczony przykład ilustrujący nieredagowaną wypowiedź mężczyzny (lat 56, zam. w Puńsku, wykształcenie zawodowe), zarejestrowaną podczas oficjalnego spotkania wraz z jej polskim wariantem w tłumaczeniu kobiety (lat 50, zam. w okolicach Puńska, wykształcenie średnie licealne).

Tabela 6. Przykład tekstu nieredagowanego z zasobów podkorpusu C. Nagranie z roku 2006.

Lp.	Wersja litewskojęzyczna	Wersja polskojęzyczna
1.	[M56] Mislinu, labai sunku kalbėc po tokio koncerto, po tokių kūrinijų ir esam visų pirma sužavėti geram darbui. [przerwa na tłumaczenie] Ir aš manau, kad šios dainos visos buvo skirtos mūsų Dzūkijai. [przerwa na tłumaczenie] Aš norėtau savo vardu tikrai jums pasakyc ačū.	[K50] Wydaje mi się, że jest trudno rozmawiać po takim koncercie, po takich utworach przede wszystkim nabieramy sił do dobrej pracy. I ja myślę, że te wszystkie piosenki były dedykowane dla nas. Ja w swoim imieniu na prawdę chciałbym Państwu podziękować.

Zasoby podkorpusu C są zrównoleglone na poziomie zdania i zapisane w plikach z rozszerzeniem TMX. Przykładowy efekt zrównoleglenia zostaje przedstawiony w tabeli 7. Pliki w formacie TMX zostają załadowane do programu obsługującego korpus.

Tabela 7. Podkorpus C. Przykład pliku włączonego do korpusu w oryginalnej postaci.

```
<tmx version="1.4">
<header creationtool="NOVA Text Aligner" datatype="plaintext" segtype="sentence" o-tmf="ATM" srclang="lt" adminlang="lt" creationtoolversion="1">
</header>
<body>
<tu>
<tuv xml:lang="lt"><seg>Mislinu, labai sunku kalbėc po tokio koncerto, po tokių kūrinijų ir esam visų pirma sužavėti geram darbui.
```

```

</seg></tuv>
<tuv xml:lang="pl"><seg>Wydaje mi się, że jest trudno rozmawiać po takim
koncercie, po takich utworach przede wszystkim nabieramy sił do dobrej pracy.
</seg></tuv>
</tu>
<tu>
<tuv xml:lang="lt"><seg>Ir aš manau, kad šios dainos visos buvo skirtos mūsų
Dzūkijai.
</seg></tuv>
<tuv xml:lang="pl"><seg>I ja myślę, że te wszystkie piosenki były dedykowane
dla nas.
</seg></tuv>
</tu>
<tu>
<tuv xml:lang="lt"><seg>Aš norėtau savo vardu tikrai jums pasakyc ačū.
</seg></tuv>
<tuv xml:lang="pl"><seg>Ja w swoim imieniu na prawdę chciałbym Państwu
podziękować.
</seg></tuv>
</tu>
</body>
</tmx>

```

Zrównoleglenie polskich i litewskich tekstów pozwala na jednoczesną obsługę tekstów w dwóch językach. Na przykład, po wpisaniu w oknie wyszukiwarki formy *piosenka*, z podawanego w tabeli 6. jako przykładowego tekstu zostanie wyświetlone polskie zdanie *I ja myślę, że te wszystkie piosenki były dedykowane dla nas.* z podświetloną formą *piosenki*, a obok niego zostanie podane odpowiadające mu zdanie litewskie *Ir aš manau, kad šios dainos visos buvo skirtos mūsų Dzūkijai.*

3. Perspektywy rozwoju EKorpGP

Dotychczas wykonywane prace nad konstrukcją korpusu były podporządkowane prowadzonym przez autorkę badaniom nad modalnością oraz kwantyfikacją zakresową w gwarze puńskiej. Obecnie, gdy osiągnięto zaplanowany cel naukowy – ukończono monografię (por. Roszko, w druku), zakres stawianych sobie zadań w kontekście korpusu uległ zmianie. Zapa-

dła decyzja o przygotowaniu wersji korpusu do publikacji w sieci. Jednak aby korpus mógł służyć nie tylko badaczom gwary puńskiej, postanowiono, że wszystkie litewskojęzyczne zasoby zostaną przetłumaczone na język polski. Podejmując taką decyzję, kierowano się zarówno lingwistycznymi, jak i pozalingwistycznymi przesłankami. Otóż zawarty w EKorpGP materiał gwarowy pochodzi z lat 1986–2012. Jest to okres burzliwych zmian politycznych, ekonomicznych, cywilizacyjnych. Zamieszkująca północno-wschodnie tereny przygraniczne litewska mniejszość narodowa jeszcze w latach 80. minionego wieku żyła w dużej izolacji. Bacznie strzeżona granica państwowa oddalała Litwinów puńskich od Litewskiej SRR (ZSRR) i jednocześnie „oddalała” od Polski. Przez Puńsk i okolice nikt nie przejeżdżał, bo z Puńska drogi wychodziły tylko w jednym kierunku – ku centralnej Polsce. Rolnicze okolice Puńska, choć malownicze, nie były celem wycieczek ani letniego wypoczynku mieszkańców Polski. Odległe od Warszawy o 300 km nie mogły konkurować z mazurskimi jeziorami czy augustowskimi lasami. Wybitnie rolniczy Puńsk nie przyciągał też klasy robotniczej poszukującej pracy w przemyśle. Dla mieszkańców Puńska i okolic oraz samej gwary puńskiej oznaczało to pełną izolację. Jednak uwarunkowania polityczne przełomu lat 80. i 90. XX wieku zapoczątkowały lawinę przemian w miasteczku. Wystarczy przywołać takie fakty jak upadek ZSRR, powstanie Republiki Litewskiej, otwarcie granic na Wschód i Zachód, w tym wstąpienie Rzeczypospolitej Polskiej i Republiki Litewskiej do NATO, Unii Europejskiej, strefy Schengen, by zdać sobie sprawę, że panujące zwyczaje w Puńsku musiały ulec zmianie. Przemiany ekonomiczne, rozwój nowych technologii czy doświadczany ostatnio kryzys gospodarczy również nie były bez wpływu na Puńsk i okolice. Bez wątpienia wszystkie wyżej wymienione czynniki pozostawiały ślad w systematycznie rejestrowanym materiale. Po pierwsze, dla językoznawców zasoby korpusowe odzwierciedlają dynamikę zachodzących zmian w samej gwarze. Bez trudu bowiem można zaobserwować jednoczesny wpływ na gwarę obu języków normowanych – polskiego i litewskiego. Po drugie, dla innych badaczy – korpus to bogate źródło informacji o charakterze socjologicznym, kulturoznawczym, etnograficznym, folklorystycznym, historycznym. Nie wyklucza się również zainteresowania korpusem osób związanych z polityką (międzynarodową – regionalną, też wewnętrzną, kształtującą politykę państwa wobec mniejszości narodowych), por. fragment wypowiedzi zamieszczony w tabeli 8. (dot. bitwy pod Grunwaldem, wypowiedź z 2010 r.).

Tabela 8. Fragment wypowiedzi kobiety, lat 75, wykształcenie policealne, zam. w okolicach Puńska.

Podkorpus A	Tłumaczenie na język polski
<p>Aš labai gerbiu visus istorikus... Mūsų internete lietuvių kalba, nežinau ar jums teko, protestavau, kad nevisiškai teisingai pagal istoriją Šapoko ir lenkų kai kurių, paduota. Kažkoks tai žiniovas, istorikas, gał mūsų krašto, gał kur, visgi sumaišė sparnus. Sako, kad tai buvo kairam sparni lietuviai.</p> <p>Romuk, aš turu Šapoką, turu planą... komunistų laikų enciklopediją tai aš ten ir matau, kad lauke buvo dešiniam sparne. Sako kairys sparnas puolė pirmiausia, kad ten kur lietuviai stovėjo, na gał jie susimaišę buvo, bet plane ir istorijoj Šapoko paduota, kad visgi dešiniam. Nežinau, ar mūsų istorikai klysta, nejaugi Šapokas, kurio istorija buvo akceptuota net komunistiniais laikais, ir kartota, leista, patobulinta, nes pirmą kartą tai išleista, turbūt trisdešimt aštuntais metais.</p>	<p>Bardzo szanuję wszystkich historyków... W naszym Internecie w języku litewskim – nie wiem, czy wam się zdarzyło – protestowałam, że podano (fakty) częściowo zgodne z zamieszczonymi w historii Šapoki i niektórych polskich (historyków). Jakiś ekspert, historyk, może i od nas, a może i nie od nas, całkiem pomylił skrzydła. Pi-sze, że Litwini byli na lewej flance. Romku, ja mam Šapokę, mam schemat (bitwy)..., encyklopedię z czasów komunistycznych, to ja tam czytam, że w polu stali na prawej flance. Podaje (wyżej wspomniany ekspert/historyk), że najpierw do walki ruszyło lewe skrzydło, tam, gdzie stali Litwini, a może oni pomieszani byli, ale na planie i w historii Šapoki podano, że jednak na prawym. Nie wiem, czy nasi historycy się myślą, czyżby Šapoka (mylił się), autorstwa którego historia była akceptowana także za czasów komunistycznych, wznawiana, drukowana, uzupełniana, bo po raz pierwszy została wydana chyba w trzydziestym ósmym roku.</p>

3.1. O dwóch polskich wariantach podkorpusu A. Podczas prac przygotowawczych do spolszczenia zasobów podkorpusu A okazało się, że możliwe są dwa warianty językowe. Pierwszy z nich byłby przekładem wykonanym przez mieszkańców Puńska i okolic, drugi zaś – przez zawodowych tłumaczy. Ten właśnie pomysł zrodził się podczas analizy językowej pierwszych tłumaczeń wykonanych przez Litwinów puńskich. Uważa się, że wprowadzenie dwóch wariantów polskiego tłumaczenia może zainteresować dialektologów polskich, badających obszar Suwalszczyzny. Prawdopodobnie – również badaczy zajmujących się polszczyzną na wschód od polskiej granicy. Z pewnością polski (miejscowy) wariant przekładu może być podstawą do badań nad procesami interferencji w mowie dwujęzycznych Litwinów puńskich. Procesy interferencji

językowej w mowie dwujęzycznej społeczności litewskiej z Puńska i okolic na Suwalszczyźnie były już przedmiotem badań prowadzonych przez N. Birgiel (2002). Niewykluczone, że dane korpusowe mogłyby wpłynąć na uszczegółowienie tego zagadnienia. W tabeli 9. zamieszczono przykład dwóch wariantów polskiego tłumaczenia.

Tabela 9. Podkorpus A. Tekst gwarowy i dwa warianty tłumaczeń na język polski.

Po litewsku	I wariant polski	II wariant polski
[M80] Tai sveikas, tep seniai matėmės.	Witaj, tak dawno widzieliśmy się.	No, cześć! Ileż to się nie widzieliśmy!
[M51] Man penkiasdešimt, jums... aštuoniasdešimt.	Mam pięćdziesiąt, a wam osiemdziesiąt.	Ja mam już pięćdziesiąt lat, a Pan (chyba) z osiemdziesiąt.
[M80] Man aštuoniasdešimt, jau peržangiau, jau balandžo mėnesį, jau devintą dešimtą pradedu. O kap tėvukas, kiek metų?	Mam osiemdziesiąt, już przekroczyłem w miesiącu kwietniu, już zaczynam dziewiątą dziesiątkę. A jak tata, ile ma lat?	Mam skończoną osiemdziesiątkę, w kwietniu rozpocząłem już dziewiątą. A jak miewa się tato, ile ma lat?
[M51] Septyniasdešimt septynerių.	Siedemdziesiąt siedem.	Siedemdziesiąt siedem.
[M80] Aha, tai da už mani jaunesnis.	Aha, to jest jeszcze ode mnie młodszy.	Aha, to jest młodszy ode mnie.
[M51] Jaunesnis, taip.	Młodszy, tak.	Tak, młodszy.
[M80] Prašom sėskit biskį. Tai kas tep pas jums, kap gyvenasi.	Proszę, usiądźcie trochę. To jak tam u was, jak się żyje.	Proszę, siądźcie choć na chwilę. Co u was słychać?
[M51] A, kažkap ten gyvenam. Vaikai dideli jau, galima sakyt. S-ai devyniolika metų. Jau studijuot važiuos į Gdanską.	A, jakoś tam żyjemy. Dzieci są już, można powiedzieć, duże. S. ma dziesięćnaście lat. Już pojedzie na studia do Gdańska.	Tak, jakoś żyjemy. Dzieci, można powiedzieć, wydorosłały. S. – dziesięćnaście lat. Zraz wyjedzie do Gdańska na studia.
[M80] Jau mūs studijas jau beigia beveik kap katriejį. Katriej turėjo galimybę – ėjo, o kici da, da S-o tai vienas da paauglys.	Nasi już co niektórzy kończą studia. Którzy mieli możliwość – to poszli, a drudzy jeszcze, S-a to jeden jeszcze podrostek.	Niektóre nasze (wnuki) już kończą studia. Te, które miały taką możliwość, to poszły, a inne jeszcze, jak S-a, to młodzieńcy.

Po litewsku	I wariant polski	II wariant polski
[K47] Mes tai tetų tai buvom susicikį. Važavo in kapines.	My ciocię to spotkaliśmy. Jechała na cmentarz.	Spotkaliśmy ciocię. Jechała na cmentarz.
[M80] Jį sakė, kad buvo susicikį. Nežinoj, kad jūs ateisit. Va, neseniai išėjo. ...kokios penkios mažum minutos kap... Tai jos kaimynka prikalbino, kad aitų.	Ona mówiła, że spotkała. Nie wiedziała, że wy przyjdziecie. Ot, niedawno wyszła. ...może gdzieś pięć minut temu, jak... To jej sąsiadka namówiła, aby poszła.	Mówiła, że was spotkała. Nie wiedziała, że przyjdziecie. I całkiem niedawno wyszła, jakieś pięć minut temu, jak... sąsiadka namówiła ją, by wyszła.
[K47] Nuėjom mes pažūrėc in tą skanseną tenai, buvo parodos kokios atsidarymo, paskui tuos namukus aplankėm, pažiūrėjom kas yra tuose namukuose ir sakom ir važuojam, ca vesailia nuvažiavo su arkliais ir motociklų ciek daug.	Poszliśmy my do tego skansenu zobaczyć, było otwarcie jakiejś wystawy, potem zwiedziliśmy te domki, zobaczyliśmy, co jest w tych domkach i mówimy – jedziemy, tu wesele pojechało konno i motocykli tak dużo.	Poszliśmy zobaczyć skansen. Tam było otwarcie jakiejś wystawy. Później zwiedziliśmy chałupy, zobaczyliśmy, co w nich jest eksponowane, i mówimy, że trzeba jechać. Tutaj goście weselni powozem konnym pojechali i było dużo motocykli.
[M80] O jau kedaisi kap jau mašinu, tai jau buvo gerai jau, o dar jau dar arkliais gerai... tai jau turi keletų porų ir tas bričkėlas turi ir pavežėja, kap jau kas akvacina.	A kiedyś to już jak był samochód, to było dobrze już, a teraz dobrze konno... to już ma kilka par i te bryczki ma i podwiezie, jak już ktoś chce.	A kiedyś jak jechali samochodem, to było coś, a teraz jeszcze lepiej końmi... Ma kilka par (koni) i bryczki, to i podwozi, jak ktoś poprosi.

Legenda:

M80: Mężczyzna, lat 80, rolnik, wykształcenie podstawowe, zam. w Puńsku (nagranie z 1998 r.)

K47: Kobieta, lat 47, wykształcenie średnie licealne, zam. w okolicach Puńska

M51: Mężczyzna, lat 51, wykształcenie wyższe, zam. w centralnej Polsce

3.2. Wybrane cechy polszczyzny w mowie Litwinów puńskich. W podkorpusie B do osobnej analizy zostały wybrane wypowiedzi Litwinów puńskich, którzy zadeklarowali swój pierwszy kontakt z językiem polskim na okres przypadający nie wcześniej niż początek nauki w szkole podstawowej. Jak się okazuje, tylko przedstawiciele pokoleń średniego i starszego spełniają ten warunek.

Wydaje się, że badania porównawcze polszczyzny w wersji Litwinów (którzy dorastając, posługiwali się tylko językiem litewskim, a dopiero w późniejszym okresie opanowali język polski jako drugi) z gwarami polskimi na Suwalszczyźnie oraz na wschód od granicy polskiej (np. na Litwie) mogą pomóc w ustaleniu, co determinuje typowe w polszczyźnie Litwinów puńskich naleciałości. Proponowane porównanie może również być pomocne w określeniu pochodzenia cech osobliwych polszczyzny regionalnej na Litwie czy Białorusi.

W wymienionych poniżej przykładach zostają zaakcentowane cechy polszczyzny Litwinów puńskich, które uzyskują najwyższy wskaźnik powtarzalności w oparciu o wyselekcjonowany w ramach podkorpusu B materiał (objętość blisko 96 000 tokenów).

Największą trudność sprawia Litwinom puńskim właściwy dobór formy aspektowej, np.:

- [1] *On wiesił się.* zamiast spodziewanego w danym kontekście *On powiesił się.*
- [2] *Dostał pozwolenie do zakładania firmy.* zamiast *Dostał pozwolenie na założenie firmy.*
- [3] *Gdzie robione?* zamiast *Gdzie wyprodukowane?* itd.

W oparciu o analizę materiału można odnieść wrażenie, że częściej zostaje wybrana forma niedokonana. W przykładzie [1] niepoprawną formę *wiesił się* należy uznać za reprezentującą aspekt dokonany w kontraście do *wieszał się*. W przykładzie [2] aspekt uwidacznia się w formie odczasownikowej *zakładanie*.

Innym typowym problemem, przed jakim stają Litwini puńscy, jest dylemat, czy użyć wyrażenia bez przyimka, czy też z przyimkiem. W ostatnim wypadku dochodzi również kwestia, jakiego użyć przyimka. W tym miejscu należy nadmienić, że gwarę cechuje wysoki współczynnik bezprzyimkowego użycia przypadków, por. (Roszko, w druku):

- [4] *Daj dla mnie.* zamiast *Daj mi.*
- [5] *Przyjechał z samochodem* zamiast *Przyjechał samochodem.*
- [6] *Służy dla ojczyzny.* (żołnierz) zamiast *Służył ojczyźnie.*
- [7] *Zadzwoń mi!* zamiast *Zadzwoń do mnie!* itd.

Analiza przykładów typu [4]-[7] wskazuje, że w przekonaniu Litwinów puńskich typową cechą języka polskiego jest duża liczba przyimków. W zdaniach [4]-[6] typowe gwarowe konstrukcje bezprzyimkowe (*duok* + celownik ; *atvažiavo* + narzędnik ; *tarnavo* + celownik) zostają błędnie oddane za pomocą polskich wyrażań zawierających przyimki. W zdaniu [7] odnotowuje się przypadek odwrotnie proporcjonalny do wcześniejszych. Tym razem oczekiwana prawidłowa polska konstrukcja z przyimkiem zostaje zastąpiona bezprzyimkowym użyciem celownika, identycznie jak w gwarze, por. gwar. *paskambink* + celownik. W oparciu

o materiał zgromadzony w podkorpusie B można wysnuć wniosek, że intuicja Litwinów puńskich co do użycia przyimka zawodzi. Gdy w gwarze i w języku polskim obserwuje się analogiczne użycia wyrażen bezprzyimkowych, to w polszczyźnie w wydaniu puńszczan odnotowuje się wyrażenia przyimkowe. Gdy gwarowemu wyrażeniu bezprzyimkowemu odpowiada polska konstrukcja przyimkowa, to Litwini puńscy stosują w języku polskim wyrażenie bezprzyimkowe.

W polszczyźnie Litwinów puńskich charakterystyczne są formy czasów dokonanych, por.:

- [8] *On był wyjechałszy do Warszawy, jak my wrócili do domu.* zamiast *On już wyjechał, kiedy my wróciliśmy do domu.*

Pewną trudność Litwinom puńskim sprawia wyrażenie miejsca i kierunku, por.:

- [9] *Masz na siebie skarpety?* zamiast *Masz na sobie (stopach) skarpety?*
[10] *Nie wchodzi w nich* (w spodnie). zamiast (*) *Nie mogę w nie wejść.*

Jest to warunkowane częstym użyciem w gwarze przypadków bezprzyimkowych dla wyrażenia miejsca, kierunku oraz znajdowania się czegoś w czymś drugim. W przykładzie [10] ujawnia się jeszcze inna cecha przenoszona z gwary. Jest nią mianowicie użycie form czasu teraźniejszego dla wyrażenia modalności, por. też inny przykład:

- [11] *Nie znajduje pieniędzy (i-ów).* zamiast *Nie mogę znaleźć pieniędzy.*

W polszczyźnie Litwinów puńskich wyraźnie różnicuje się znaczenia ilościowe. Gdy mowa jest o wszystkich elementach danego zbioru, wówczas znajduje zastosowanie biernik, w przeciwnym razie – dopełniacz, por.:

- [12] *Kupiła tanio jajek.* zamiast spodziewanego wariantu *Kupiła tanio jajka.*

Do grupy typowych (o wysokiej frekwencji) cech polszczyzny w wydaniu Litwinów puńskich można zaliczyć mylenie przypadków, np.:

- [13] *Boli mi głowe* (biernik) zamiast *Boli mnie głowa.*
[14] *Swendzi mi cało skóre* (jw.) zamiast *Swędzi (mnie) cała skóra / Wszędzie mnie swędzi.*
[15] *Zapłace ratami* zamiast *Zapłacę w ratach.*
[16] *Jeszcze jest ich.* zamiast *Jeszcze tu są.* (w znaczeniu: jeszcze nie poszli sobie).

W polszczyźnie puńszczan zdarzają się kalki językowe. Oto niektóre z za-notowanych przypadków, np.:

- [17] *Buty obgryzają mi nogi.* zamiast *Buty obcierają mi nogi.*
[18] *Musze przebrać łóżko.* zamiast *Muszę zmienić pościel.*

Por. też przykład [3], w którym forma *robić* odpowiada polskiej *produkować*.

Przedstawione tu przykłady typowych błędów językowych zostały opracowane w oparciu o niewielki materiał podkorpusu B (o objętości blisko 96 000 tokenów). Dalsza rozbudowa korpusu z pewnością pozwoli na bardziej szczegółową analizę zjawiska.

4. Podsumowanie

Włączone do EKorpGP zasoby językowe przedstawiają dużą wartość poznawczą dla polskich i litewskich dialektologów. Szczególnie cenny dla badaczy gwar litewskich, w tym gwary puńskiej, jest podkorpus A. Bezsprzecznie, gwara ewoluuje. Typowe cechy gwarowe zanikają. Wraz z nimi stopniowo obumiera gwara. Młodsze pokolenie dobrze posługuje się litewskim językiem normowanym. Średnie pokolenie, w zależności od wykształcenia, w mniejszym lub większym stopniu utrzymuje cechy gwarowe. Wypowiedzi przedstawicieli pokolenia starszego zachowują najwięcej charakterystycznych cech gwarowych, choć również w ich mowie obserwuje się wpływy litewskiego języka normowanego, szczególnie w zakresie leksyki.

Podkorpusy B i C stanowią dobrą podstawę do badań nad procesami interferencji językowej w mowie dwujęzycznych Litwinów puńskich. Potencjalni odbiorcy tych podkorpusów to również dialektolodzy zajmujący się gwarami Suwalszczyzny, Podlasia, a także badacze polszczyzny regionalnej na wschód od granicy polskiej.

Proponowane „spolszczenie” EKorpGP (dotyczy podkorpusu A) wraz z podkorpusami B i C – to wiarygodne źródło informacji dla szerokiego kręgu badaczy zjawisk natury socjologicznej, etnologicznej, kulturoznawczej, historycznej. Niewykluczone, że poruszana w wypowiedziach Litwinów puńskich tematyka zainteresuje osoby kształtujące politykę wobec mniejszości narodowych w Polsce.

Nieprzeciętna wartość zgromadzonych w EKorpGP materiałów wynika z faktu, że są one odzwierciedleniem przemian politycznych, gospodarczych, cywilizacyjnych okresu burzliwych przeobrażeń w Polsce, na Litwie i w Europie (lata 1986–2012).

Literatura

- Birgiel N., 2002, *Procesy interferencyjne w mowie dwujęzycznej społeczności litewskiej z Puńska i okolic na Suwalszczyźnie*, Punsko „Aušros leidykla“, Warszawa – Puńsk.
- Krištopaitė D. (red.), 1998, *Nuo Punsko iki Seinų. Iš Juozo Vainos rinkimų*. II. Punsko „Aušros“ leidykla, s. 282.

- Krištopaitė D. (red.), 1999, *Nuo Punsko iki Seinų. Iš Juozo Vainos tautosakos rinkimų*. I. Punsko „Aušros“ leidykla, s. 317.
- Roszko D., 2013, *Experimental corpus of the Lithuanian local dialect of Puńsk in Poland. Examples of the lexical and semantic annotation*, „Cognitive Studies | Études Cognitives” 13, s. 79–95; DOI: 10.11649/cs.2013.005.
- Roszko D., (w druku), *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (na tle literackich języków polskiego i litewskiego)*, Warszawa
- Smoczyński W., 1984a, *Szkic morfologiczny litewskiej gwary puńskiej*, „Acta Baltico-Slavica” 16, s. 235–261.
- Smoczyński W., 1984b, *Zapóżyczenia słowiańskie w litewskiej gwarze puńskiej*, [w:] *Studia nad gwarami Białostockizny. Morfologia i słownictwo*, red. Maryniakowa I., Smułkowska E., [„Prace Białostockiego Towarzystwa Naukowego” 27], Warszawa, s. 179–222.
- Smoczyński W., 1986a, *System fonologiczny litewskiej gwary puńskiej*, „Acta Baltico-Slavica” 17, s. 369–385.
- Smoczyński W., 1986b, *Zapóżyczenia niemieckie w gwarze litewskiej okolic Puńska na Suwalszczyźnie*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego, Prace Językoznawcze” 82, s. 35–45.

Experimental corpus of the Lithuanian local dialect of Puńsk in Poland (a tool for the study on the Polish and Lithuanian local dialects in the Suwałki Region)

(Summary)

The linguistic resources included in EkorpGP present a great cognitive value for Polish and Lithuanian dialectologists. Particularly valuable for researchers of the Lithuanian local dialects, including the local dialect of Puńsk, is subcorpus A. The material included in EKorpGP confirms that the local dialect is incessantly evolving. The typical dialectal features vanish for the sake of the standardized Lithuanian language.

Whereas, subcorpora B and C are a good base for studies on the processes of the language interference in the speech of the bilingual Lithuanians of Puńsk. Amongst the potential recipients of these subcorpora it is possible also to find dialectologists dealing with the local dialects of the Suwałki Region, as well as researchers of the Polish language spoken to the east of the Polish border.

The polonization suggested in EKorpGP (with reference to subcorpus A) along with subcorpora B and C will ensure a credible source of information for a wide circle of researchers of the phenomena of sociological, ethnological, cultural and historical nature. It should also be taken into account that the issues raised in the utterances given by the Lithuanians of Puńsk will capture the interest of politicians dealing with the national minorities in Poland.

The extraordinary value of the materials collected in EKorpGP results from the fact that they reflect the economic, political and civilization transformations of the turbulent period of 1986–2012 in Poland, Lithuania and Europe.