

<https://doi.org/10.19195/0137-1150.175.2>

Data przesłania artykułu: 29.11.2020

Data akceptacji artykułu: 25.06.2021

DANUTA ROSZKO

Uniwersytet Warszawski, Polska
(University of Warsaw, Poland)

ROMAN ROSZKO

Instytut Sławistyki Polskiej Akademii Nauk, Warszawa, Polska
(Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland)

Projekcja znaczeń w wielojęzycznych korpusach wraz z przykładami jej zastosowania w badaniach korpusowych nad językiem polskim

The projection of meanings in multilingual corpora and examples of its application in corpus research on the Polish language

Abstract

The Institute of Slavic Studies of the Polish Academy of Sciences, as part of the CLARIN-PL consortium, is engaged in the creation of multilingual corpora of Slavic and Baltic languages. Some of the planned corpora have already been made available in the *dSpace* CLARIN-PL repository and in the *KonText* online browser. One of these corpora is the Polish–Lithuanian corpus, which in this article serves to illustrate the possibilities of applying CLARIN-PL corpora in research into the Polish language. The Lithuanian language has an archaic structure in comparison with Polish. Lexemes are characterised by a clearer structure. There are regular connections between grammatical form and meaning, with a small number of exceptions. These features of Lithuanian ensure that the projection of strongly formalised meanings onto Polish-language resources is a straightforward operation, leading to a significant narrowing of Polish-language resources, which can be analysed in the context of unequivocal interpretations based on Lithuanian meanings.

Keywords: parallel CLARIN-PL corpora, meaning projection, Lithuanian language, Polish language

Проекция значений в многоязычных корпусах и примеры ее применения в корпусных исследованиях польского языка

Аннотация

Институт славистики Польской академии наук совместно с CLARIN-PL занимается созданием многоязычных корпусов для славянских и балтийских языков. Часть запланированных корпусов уже доступна в репозитории *dSpace* CLARIN-PL и в онлайн браузере *KonText*. Один из таких корпусов — это польско-литовский корпус, который в данной статье послужит источником иллюстрации возможностей использования корпусов CLARIN-PL в исследованиях польского языка. Литовский язык имеет сохранившуюся архаичную структуру. Лексемы характеризует более прозрачное строение. Между грамматической формой и значением существует более регулярная связь. Исключения крайне редки. Именно эти отличительные особенности литовского языка приводят к тому, что проекция сильно формализованных значений на польскоязычные ресурсы является простой процедурой, ведущей к существенному сужению польских ресурсов, которые могут анализироваться в контексте однозначно воспринимаемых на фоне литовского языка значений.

Ключевые слова: параллельные корпуса CLARIN-PL, проекция значений, литовский язык, польский язык

Wprowadzenie

Już w XIX wieku twórcy gramatyki historyczno-porównawczej zauważyli, że języki z sobą spokrewnione w większym stopniu łączy płaszczyzna znaczeniowa niż samo podobieństwo form. Aby ustalić pokrewieństwo między językami, opisywali prawa głosowe, których działanie doprowadzało w określonym miejscu i czasie do różnicowania się powłoki fonetycznej języków wywodzących się z tego samego prajęzyka. Prawa głosowe determinowały dyferencjację języków. Jednak, co można zauważyć, niektóre języki, grupy czy rodziny językowe dokonały tych zmian zdecydowanie mniej. Językiem, w którym takich zmian zaszło definitywnie mniej, jest język litewski, natomiast takim, w którym były one liczne, jest z pewnością język polski. Zważywszy na wskazany odmienny charakter języków polskiego i litewskiego, mamy prawo sądzić, że bardziej stabilne na przestrzeni wieków litewskie struktury formalne mogą współcześnie wykazywać wyższy i jednocześnie nieskomplikowany stopień relacji zachodzących między formą a jej funkcją. Zamierzamy zatem ustalić, czy ta właśnie zależność (forma–funkcja; typowa dla języka litewskiego) może okazać się przydatna w badaniach nad językiem polskim. Najprostszym i najbardziej efektywnym sposobem zastosowania litewskich faktów gramatycznych, jak myślimy, jest rzutowanie jednoznacznie odczytywanych w strukturach języka litewskiego (to jest na płaszczyźnie formalnej) znaczeń na zasoby polskie, ściśle skorelowane z litewskimi w korpusach

dwu- i wielojęzycznych. Do testów wybraliśmy najnowsze dwujęzyczne korpusy polsko-litewski i polsko-rosyjski z serii *Corpora of Baltic and Slavic Languages* (Duszkin, Roszko, Roszko, 2021), których zasoby są ręcznie zrównoleglone oraz automatycznie lematyzowane i tagowane (warstwa fleksyjna i składniowa). Konstrukcją dziesięciu korpusów z przedstawionej serii zajmuje się zespół językoznawców z Instytutu Sławistyki PAN w ramach zadań konsorcjum CLARIN-PL¹.

1. Zachowawczy charakter języka litewskiego a innowacyjna natura języka polskiego

Naszym celem nie jest systematyczne dokumentowanie zmian historycznych obu języków. Chcemy jednak zwrócić uwagę, że zmian, które mogą prowadzić do zacierania się granic między morfemami, jest mniej w języku litewskim niż w polskim. Zauważmy chociażby dysproporcje ilościowe na płaszczyźnie fonetycznej obserwowane w zestawieniu litewskiego leksemu *sūnus* [sūnūs] ‘syn’ i jego polskiego odpowiednika *syn* (< *syñ* < *syñ* < **sūnu* < **sūnus*). Rekonstruowana praindoeuropejska postać tego leksemu to **suH-nu-* (masc.), wywodzona z pierwiastka praindoeuropejskiego **seuH-/*suH-* ‘rodzić, wydawać na świat’ (Smoczyński; 2007, 2019, s. 1563). Analogiczne dysproporcje dostrzeżemy w innej parze leksemów: litewskiego *sapnas* ‘sen’ i polskiego *sen* (< *s̃ñ* < *s̃ñ* < *s̃ñ* < **s̃p-ñs* < **sap-nas*). Rekonstruowana praindoeuropejska postać tego leksemu to **suép-no-* (Smoczyński, 2007; 2019, s. 1355). Mniej zmian fonetycznych w formach litewskich to bardziej czytelna struktura słowoform. Zacieranie granic między morfemami, w tym powstawanie amalgamatów, może obniżać potencję poszczególnych formantów. Dlatego, jak uważamy, mniejsze zmiany na płaszczyźnie formalnej sprzyjają regularności i żywotności urabiania poszczególnych form w języku litewskim. Przyjrzyjmy się kilku przykładom tej regularności w konfrontacji polsko-litewskiej.

Renata Grzegorzczakowa zauważa, że tworzenie form deminutywnych w języku polskim jest z reguły zabiegiem powszechnym i regularnym (1979, s. 53), jednak gdy zestawimy skuteczność tworzenia deminutywów w językach polskim i litewskim, od razu dostrzeżemy, że ten proces słowotwórczy przebiega z mniejszymi oporami właśnie w języku litewskim. Zauważmy, że o ile w języku polskim urobienie deminutywów od takich form, jak: *Litwa*, *młodość*, *kierowca*, *romantyk*, jest kłopotliwe, o tyle w języku litewskim jest nie tylko możliwe, lecz również pozbawione negatywnych konotacji; por. deminutywa: *Lietuv-ėl-ė* (< *Lietuv-a* ‘Litwa’), *jaunyst-ėl-ė* (< *jaunyst-ė* ‘młodość’), *vairuotoj-ėl-is* (masc.) (< *vairuotoj-as* ‘kierowca (masc.)’), *vairuotoj-ėl-ė* (fem.) (< *vairuotoj-a* ‘kierowca (fem.)’), *romantik-ėl-is* (< *romantik-as* ‘romantyk’). Innym podobnym do przedstawionego

¹ <http://CLARIN-PL.eu/>

zjawiskiem jest urabianie fminatywów w tych językach. Znow jesteśmy skłonni twierdzić, że historycznie mniejsze zmiany fonetyczne w języku litewskim zapewniają pełny automatyzm tworzenia form fminatywnych; por. *kunig-as* ‘ksiądz’ i *kunig-ė* ‘ksiądz (fem.)’, *siūvik-as* ‘szewc’ i *siūvik-ė* ‘szewc (fem.)’ itp.

Testem formalnym przejrzystości struktur polskich i litewskich jest konstrukcja zapytań z użyciem wyrażeń regularnych w języku CQL (język zapytań korpusowych; por. przykłady takich zapytań w sekcji 2). Utworzenie zapytań dla języka litewskiego zazwyczaj nie wymaga konstrukcji skomplikowanych, wielowarunkowych formuł. Analogiczne w treści zapytania dla języka polskiego niestety składają się z bardzo wielu warunków, w tym szeregu wyjątków i wykluczeń. Twórcy wyszukiwarek korpusowych ograniczają jednak długość pola zapytania. Oznacza to, że w przypadku języka polskiego konieczne będzie stopniowe zawężanie uzyskanych wyników w kolejnych etapach. Niestety większość wyszukiwarek nie zapewnia pełnej funkcjonalności wynikającej z użycia składni zapytań przy już wyświetlonej odpowiedzi na pierwsze zapytanie.

Reasumując, przypuszczamy, iż łatwość tworzenia nowych słowoform, leksemów, derywatów, któremu na przeszkodzie nie stoją liczne ograniczenia właściwe płaszczyźnie formalnej, może sprzyjać wyższej regularności kodowania treści w samych formach językowych. Z tej przesłanki wynika kolejne założenie, mówiące o tym, że języki o bardziej przejrzystej strukturze płaszczyzny formalnej mogą być w istocie bardziej konsekwentne i rygorystyczne zarówno w kodowaniu znaczeń, jak i w ich wyrażaniu. Tę właśnie cechę przypisujemy językowi litewskiemu i zamierzamy ją zastosować w automatycznym wskazywaniu form/wyrażeń/kontekstów polskich w procesie projekcji znaczeń z języka litewskiego na polski w korpusach dwujęzycznych — polsko-litewskich.

2. Polish-Lithuanian Parallel Corpus „2”

W przestrzeni wirtualnej funkcjonuje wiele korpusów wielojęzycznych, które zawierają zrównoległone zasoby polsko-litewskie. Można tu wymienić portal Linguee², obsługujący wyszukiwanie w interesującej nas parze języków, jednak tylko w ograniczeniu do poszczególnych leksemów (słów bazowych) i fraz. Niestety w tym korpusie nie jest możliwe uzyskanie wyniku, który zawierałby serię wielu różnych leksemów spełniających dowolne kryterium (na przykład wszystkie słowoformy będące rzeczownikami rodzaju męskiego). Przykładem bardziej rozbudowanego funkcjonalnie korpusu jest InterCorp (Čermák, Rosen, 2012), którego składnia zapytań choć nie obejmuje wyszukiwania według lematów (słów bazowych) i parametrów morfosyntaktycznych w języku litewskim, to umożliwia przeszukiwanie tych zasobów z zastosowaniem składni wyrażeń regularnych; por. przykład (1).

² <https://www.linguee.com>.

(1) [word=".*ël(is|io|iui|ï|iu|y|je|yj|iai|iü|iams|ius|iais|iuose|iuos)"
word=".*ël(è|ès|ei|ë|e|ëj|ëje|iü|ëms|es|ëmis|ése)"]³

Przywołane zapytanie (1) wyszukuje wszystkie litewskie formy (słowofor-my), które kończą się na: *-ëlis*, *-ëlio*, *-ëliui*, *-ëlj* itd. Łącznie w odpowiedzi uzyskuje się dla dowolnego leksemu potencjalny szereg 26 różnych form deklinacyjnych, składających się na cały paradygmat form męskich i żeńskich w obu liczbach, które zawierają element *-ël-* (potencjalny sufiks) bezpośrednio przed fleksją; por. *pinig-as* ‘pieniądz’ i *pinig-ël-is* ‘pieniązek’. W korpusie InterCorp w odpowiedzi na tak postawione zapytanie (1) badacz otrzyma interesujące go formy (to jest zawierające wspomniany sufiks *-ël-*), ale również szereg innych form, które tylko pozornie zawierają ten sufiks, na przykład *lëlë* ‘lala’ (tu *ël* jest częścią tematu: *lël-ë*), *këlë* (czasownik) ‘podniósł | podniosła | podniosło | podnieśli | podniosły’ (także tu *ël* jest częścią tematu: *kël-ë*) czy *dëlei* (przyimek) ‘z powodu’. Brak wprowadzonej lematyzacji i anotacji morfosyntaktycznej polsko-litewskich zasobów w korpusie InterCorp⁴ nie pozwala na zawężenie wyników, na przykład do rzeczowników.

W odróżnieniu od InterCorpu moduł wyszukiwania wielojęzycznych korpusów CLARIN-PL uwzględnia lematy i anotację morfosyntaktyczną dla pary polsko-litewskiej. Dzięki temu możliwe jest uproszczenie składni zapytań oraz wprowadzenie dodatkowych warunków doprecyzowujących wyszukiwanie. W korpusie CLARIN-PL ekwiwalentem zapytania (1) z korpusu InterCorp będzie na przykład zapis:

(2) [lexem=".*ël(is|è)"]

Element <lexem> wraz z wyrażeniem regularnym <.*ël(is|è)> definiuje nam wszystkie leksemy, które w formie bazowej (słownikowej) kończą się na *-ëlis* (formy rodzaju męskiego) lub *-ëlë* (formy rodzaju żeńskiego). Przedstawiony zapis (2) można zmodyfikować tak, by w odpowiedzi na zapytanie zostały wyświetlone tylko rzeczowniki:

(3) [lexem=".*ël(is|è)" & tag="dkt.*"]

Element zawężający wyniki wyszukiwania znajduje się po spójniku „i” <&>, a mianowicie: <tag="dkt.*">, gdzie parametr <tag> (przywołujący jednostki morfosyntaktyczne) zawiera wielkość <dkt>, definiującą rzeczowniki, oraz wyrażenie regularne <.*> — zastępujące wszystkie pomijalne w wyszukiwaniu dodatkowe parametry rzeczownika (takie jak przypadek, liczba, rodzaj i inne).

Podaną w przykładzie (3) składnię zapytania można z kolei zawęzić tak, by w wynikach nie pojawiały się nazwy własne:

(4) [lexem=".* ël(is|è)" & tag="dkt.*" & tag!="tkr.*"]

³ Sformułowanie ekwiwalentnego zapytania dla języka polskiego jest wielokrotnie trudniejsze.

⁴ To ograniczenie nie dotyczy wszystkich zasobów InterCorp.

Elementem składni zapytania, który zawęży wyszukiwanie rzeczowników do nazw pospolitych, jest zapis `<tag!="tkr.*">`, w którym widzimy zanegowany parametr `<tag!>`, zawierający wielkość `<tkr>`, definiującą nazwy własne. Autorzy zastosowanej w korpusach CLARIN-PL anotacji litewskiej MorfoLema⁵ ograniczyli znakowanie tylko do nazw własnych. Dlatego w składni zapytania należało zastosować negację nazwy własnej.

Objętość wybranego do badań korpusu wynosi 16 543 470 słowoform. Zastosoby są lematyzowane i otagowane. Szerzej o tym w *Polish-Lithuanian Parallel Corpus „2”* (Roszko, Roszko, 2018), zaś o najnowszej wersji polsko-litewskiego korpusu oraz pozostałych wielojęzycznych korpusach języków słowiańskich i bałtyckich CLARIN-PL w publikacji zespołu Duszkin, Roszko i Roszko (2021).

3. Przykłady potencjalnych badań nad językiem polskim

W tej części artykułu przedstawiamy propozycje zagadnień, których analiza jest możliwa dzięki ewidentnym różnicom zachodzącym między oboma językami. Większość przykładów cytowanych w artykule pochodzi z opisanego w sekcji 2 korpusu polsko-litewskiego CLARIN-PL⁶. Poza tym opis będziemy ilustrować przykładami zaczerpniętymi z powieści Stanisława Lema *Solaris* (1961) i jej tłumaczenia na język litewski: Stanislavas Lemas, *Soliaris* (1978, przeł. Giedrė Juodvalkytė)⁷ oraz korpusu polsko-rosyjskiego CLARIN-PL.

3.1. Użycie zaimków dualnych

Odziedziczona z języka praindoeuropejskiego liczba podwójna nie zachowała się w żadnym z opisywanych języków. Utrzymane charakterystyczne dla liczby podwójnej formy są jedynie śladem wcześniej funkcjonującej kategorii. Niemniej jest coś, co zwróciło naszą uwagę podczas przeszukiwania Korpusu: mianowicie znaczna — w odniesieniu do języka polskiego — liczba użyc litewskich zaimków dualnych, por.:

(5) Pol. — *Gdzie **my** jesteśmy, Harey?*

Litew. — *Kur **mudu** esame, Hare?*

(6) Pol. — *Sartorius przekonał Snautę, że go oszukałeś. Teraz **oni** ciebie oszukają.*

Litew. — *Sartorijus įtikino Snautą, kad tu apgavai jį. Dabar **jiedu** apgaus tave.*

⁵ <http://donelaitis.vdu.lt/MorfoLema/Apie.htm>.

⁶ W dalszej części artykułu, pisząc „Korpus”, będziemy mieli na myśli polsko-litewski korpus Polish-Lithuanian Parallel Corpus „2”.

⁷ Utwory takie jak *Solaris* Stanisława Lema są częścią eksperymentalną polsko-litewskiego korpusu CLARIN-PL, lecz ze względu na brak licencji na użycie w korpusach nie są dostępne badaczom w opublikowanych na stronach CLARIN-PL korpusach języków bałtyckich i słowiańskich.

W przykładach (5)–(6) polskie zaimki mnogie *my* i *oni* korespondują z litewskimi zaimkami dualnymi *mudu* i *jiedu*. Element wyróżniony w litewskich zaimkach — *du* — to literalny odpowiednik polskiego liczebnika *dwa*. Przy czym litewskim formom *mudu*, *jiedu* należy wirtualnie przypisać znaną z Narodowego Korpusu Języka Polskiego wartość „m1” (rodzaj męski osobowy) (por. Przepiórkowski, Bańko, Górski, Lewandowska-Tomaszczyk, 2012). Litewskie odpowiedniki żeńskie to *mudvi*, *jiedvi*. Świadomość doprecyzowania liczby w języku litewskim pozwala na wyróżnienie i analizę w języku polskim kontekstów, w których informacja o dwuelementowym zbiorze może zostać pominięta. Warto też w tym kontekście zwrócić uwagę, że korelacja litewskich zaimków dualnych z polskimi zaimkami mnogimi nie jest oczywista, co zostaje uwidocznione w przykładach (7)–(9).

(7) Pol. *Ty z Hrynek i ja z Hrynek, wszyscy my z Hrynek, tylko że ja biedna sierota była, nad którą nigdy nikt litości nie miał.*

Litew. **Mudu abu** iš Meliūnų, ir aš iš Meliūnų; visi mes iš Meliūnų; tiktai aš buvau vargdienė-našlaitė, ant kurios niekas nesusimylėjo.

(8) Pol. *Pod pozorem montowania aparatury rentgenowskiej o budują anihilator pola.*

Litew. *Dėdamiesi montuoja rentgeno aparatūrą, jiedu stato lauko anihiliatorių.*

(9) Pol. *Znowu byłem w zupełnej ciemności, nie pojmowałem niczego, a przez chwilę здавало mi się, że uda mi się go przekonać i o wspólnymi siłami zaatakujemy zagadkę.*

Litew. *Aš vėl atsidūriau aklavietėje, nieko nesupratau. O jau buvau bemanęs, kad pasiseks ji įtikinti ir mudu bendrom jėgom pulsime šitą paslaptį.*

O ile w przykładzie (7) można mówić o odpowiedniości litewskiego *mudu abu* (zaimka dualnego *mudu* wzmocnionego liczebnikiem *abu* ‘obu’) i polskich zaimków *ja* i *ty* (identyfikujących mówiącego i odbiorcę, czyli dwie osoby), o tyle w (8)–(9) brakuje korelacji litewskich zaimków dualnych *jiedu* i *mudu* z jakąkolwiek polską formą. W takich przypadkach projekcja użyć zaimków dualnych w zdaniach litewskich na skorelowane z nimi segmenty w języku polskim jest już wystarczającym kryterium pozwalającym wyodrębnić konteksty do badań nad dualnością, szerzej — liczbą, w języku polskim. Opierając się na badaniach zawartych w pracy Danuty Roszko (2015), należy stwierdzić, że korelacja polskiego formalnego „zera” z litewskimi zaimkami (tu dualnymi) jest przejawem silnie rozwiniętej kategorii określoności-nieokreśloności w języku litewskim. Uogólniając wskazaną tu korelację litewskich form z polskimi „zerowymi” wykładnikami na inne analogiczne znane nam przykłady, możemy stwierdzić, że w języku litewskim w odniesieniu do języka polskiego przekazywane w wypowiedzi znaczenia zazwyczaj mają większą formalną reprezentację.

3.2. Użycie form wołacza

Analiza korpusowa wystąpień form wołacza w obu językach potwierdza stopniowy zanik użycia tych form w języku polskim. Zestawiliśmy reprezento-

wane w korpusie dzieła dziewiętnastowieczne⁸ z najnowszymi⁹. O ile w dziewiętnastowiecznych utworach polskich i ich litewskich przekładach można mówić o wysokiej zgodności w użyciu form wołacza (około 250 wystąpień na milion słowoform), o tyle w zasobach najnowszych ta korelacja zgodności została mocno uszczuplona (około 15 wystąpień na milion słowoform). Liczba odnotowanych form wołacza w języku litewskim nadal utrzymuje się na podobnie wysokim poziomie, podczas gdy liczba użyc form wołacza w języku polskim znacznie spadła. Szczegółowe wyniki tych badań zostały opisane w oddzielnym artykule (Roszko, Roszko, 2021). W przykładach (10)–(11) wskazujemy na użycie formy mianownikowej w funkcji wołacza w języku polskim.

(10) Pol. — *Gotów, **Kelvin**? — rozległo się w słuchawkach.*

— *Gotów, **Moddard** — odpowiedziałem.*

Litew. — *Pasiruošęs, **Kelvinai**? — pasigirdo ausinėse.*

— *Pasiruošęs, **Modardai**, — atsakiau.*

(11) Pol. — ***Siostra**, utopisz się.*

Litew. — *Nuskęsi, **sesute**.*

W przypadku nazw własnych użytych w tekście polskim (10) można roboczo założyć, że obce brzmienie imion może sprawiać trudność w urobieniu form przypadkowych. Jednak w powieści rejestrujemy formy, co prawda tylko dopełniacza — por. *twarz **Moddarda**, bliski głos **Moddarda**, na ekranach **Moddarda**, z tych hipotetycznych części **Kelvina**, w przypuszczeniu kolegi **Kelvina**, coś gorszego od **Kelvina*** — które świadczą, że tworzenie form przypadkowych nie jest tu trudnością czysto formalną.

Również w wypadku użycia form wołacza zauważamy, że w litewskiej wersji tekstu rejestrujemy takie jego użycia, z którymi w tekście polskim nie są skorelowane żadne leksemy. Należy odnotować, że podawane w przykładach (12)–(13) zdania litewskie pochodzą z utworów tłumaczonych z języka polskiego.

(12) Pol. *A jeszcze **o** poznawszy Swalgona...*

Litew. *O čia, **vyreli**, Svalgūnq paleido!*

(13) Pol. ***o** Pamiętają ludzie, gdy jeno pieśni nuciła, strojno chodziła, a chłopiątko w kwiatki ubierając, na rękę kołysała.*

Litew. *Ak, **broli**, žmonės pamena, kaip dainuodavo, puikiai rėdydavos, o kūdikį kaišydavo į kvietukas, supdavo lopšyje.*

Podobne do przedstawionego zjawisko zaobserwowaliśmy we współczesnych tłumaczeniach z języka rosyjskiego na polski. Z analizy rosyjskiej ścieżki dialogowej serialu fabularno-dokumentalnego *Przylapani na zdradzie* oraz jej tłumaczenia na język polski wynika, że użyciu tak zwanego nowego/współczesnego

⁸ Adam Asnyk, *Kiejstut* (1843) / Adamas Asnykas, *Kęstutis* (1897, przeł. Vincas Kudirka); Józef Kraszewski, *Kunigas* (1882) / Juozas Ignatas Kraševskis, *Kunigas* (1887, przeł. Augustinas Zeicas); Eliza Orzeszkowa, *Niziny* (1885) / Eliza Ožėškienė, *Šunadvokatis* (1901, przeł. Vincas Kalnietis, Jonas Jablonskis).

⁹ Ze względu na dużą liczbę utworów nie wymieniamy ich tytułów.

wołacza w języku rosyjskim (por. na ten temat Krongauz, 1999; Suprun, 2001; Polonskii, 2002; Anstatt, 2008) najczęściej nie odpowiada żadna forma w języku polskim; por. przykłady (14–15).

(14) Pol. **o** *Czy to nie Lusia?*

Ros. **Rebyat**, *razve eto ne Lyusiya?*

(15) Pol. **o** *Witaj! Przejdźmy się.*

Ros. **Kristin**, *zdravstvuite! Davaite proidemsysa.*

Dodawanie form w wołaczu w przekładach utworów polskich na język litewski oraz pomijanie form w wołaczu w tłumaczeniach polskich z języka rosyjskiego może być świadectwem nie tyle tendencji zanikowej wołacza jako przypadku, ile raczej okołojęzykowych faktów, na przykład ograniczania liczby bezpośrednich zwrotów do współrozmówcy.

3.3. *Ma + infinitivus* — z funktorem konieczności czy możliwości?

Rozróżnienie znaczenia polskiej konstrukcji paramorfologicznej *ma + infinitivus* może zostać automatycznie dokonane w wyniku rzutowania znaczeń z języka litewskiego. Przykłady (16)–(17) ilustrują użycie tej konstrukcji, powiązanej z pojawieniem się funktora konieczności w płaszczyźnie semantycznej zdania. Tym polskim konstrukcjom odpowiada w tekście litewskim złożenie formy osobowej czasownika *turėti* ‘mieć’ z bezokolicznikiem. Z kolei w przykładach (18)–(20), zawierających funktor możliwości w płaszczyźnie semantycznej zdania, w litewskich odpowiednikach polskiego *ma + infinitivus* nie stwierdzamy obecności formy osobowej czasownika *turėti* ‘mieć’. Wówczas polskiej konstrukcji *ma + infinitivus* zawsze odpowiada wyrażenie proste fundowane na imiesłowach czynnych. Zatem *miał zagwarantować* (por. (18)) koresponduje z litewskim imiesłowem czynnym czasu przeszłego jednorazowego *užtikrinęs* (sg. masc.) (< *užtikrinti* ‘zapewnić / zagwarantować’). W przykładzie (19) polskie *ma siedzieć* koresponduje z litewskim imiesłowem czynnym czasu przyszłego *sėdėsias* (sg. masc.) (< *sėdėti* ‘siedzieć’). W kolejnym zdaniu (20) stwierdzamy powiązanie polskiego *ma mówić* i litewskiego imiesłowu czynnego czasu teraźniejszego *sakęs* (sg. masc.) (< *sakyti* ‘mówić’). Wszystkie wyróżnione w przykładach (18)–(20) litewskie imiesłowy są wykładnikami semantycznej kategorii modalności imperceptywnej (por. Koseska-Toszeva, Korytkowska, Roszko, 2009, s. 70–195), której jednym z wykładników morfologicznych są formy trybu znanego w nomenklaturze litewskiej jako *modus relativus* (por. Ambrazas, 1990).

(16) Pol. — **Miał** tylko *powiedzieć*, że przyszedł.

Litew. — *Jis turėjo tik pasakyti, kad buvo atėjęs.*

(17) Pol. **Miała** zrobić wrażenie.

Litew. *Ji turėjo sudaryti įspūdį.*

- (18) Pol. *Miał zagwarantować nabywcy warunki nie mniej korzystne niż wskazane w zamówieniu.*
Litew. *Jis užtikrinęs pirkėją, kad sąlygos būsiąs bent jau ne blogesnės nei nurodyta užsakyme.*
- (19) Pol. *Mówił, że ma siedzieć w domu.*
Litew. *Sakėsi sėdėsiąs namuose.*
- (20) Pol. *Byli tacy dobrzy ludzie, co mu je do czytania dawali, a on czytał i zapamiętał, i teraz ma im mówić, że wszystko będzie dobrze.*
Litew. *Atsiradę tokių gerų žmonių, kurie davę jas jam pasiskaityti, ir jisai jas skaitęs ir galvon dėjėsis, ir dabar jiems sakąs, kad visa būsią gerai.*

W litewskiej części przykładów (18) i (20) odnotowujemy większą liczbę użyć form *modus relativus*. W zdaniu (18) jest nią również *būsiąs* (imiesłów czynny czasu przyszłego), która ze względu na inną konstrukcję zdania w języku polskim nie ma swojego bezpośredniego odpowiednika. Dosłowne tłumaczenie zdania litewskiego (18) można przedstawić następująco: *Miał zagwarantować nabywcy warunki, które ponoć miały być nie mniej korzystne niż wskazane w zamówieniu.* Z badawczego punktu widzenia szerszego omówienia wymaga przypadek (20). W przykładzie (20''), będącym kopią zdań z (20), w litewskiej wersji wytluszczeniem wyróżniamy wszystkie formy *modus relativus*. Podkreśleniem w polskiej wersji językowej zaznaczamy te formy, które odczytujemy jako ekwiwalenty tłumaczeniowe tychże litewskich form *modus relativus*.

(20')

Pol. *Byli tacy dobrzy ludzie, co mu je do czytania dawali, a on czytał i zapamiętał, i teraz ma im mówić, że wszystko będzie dobrze.*

Litew. *Atsiradę tokių gerų žmonių, kurie davę jas jam pasiskaityti, ir jisai jas skaitęs ir galvon dėjėsis, ir dabar jiems sakąs, kad visa būsią gerai.*

(20'')

Pol. *Byli tacy dobrzy ludzie, co mu je do czytania dawali, a on czytał i zapamiętał, i teraz ma im mówić, że wszystko będzie dobrze.*

* *Ponoć* byli tacy dobrzy ludzie, co mu je do czytania mieli dawać, a on *ponoć* czytał i *ponoć* zapamiętał, i teraz *ma* im *mówić*, że wszystko *ma* być dobrze.

Jak widać, liczba (sześciu) litewskich form *modus relativus*, które należy uznać za ewidentny wykładnik znaczeń modalności imperceptywnej, jest niewspółmiernie większa od liczby (tylko jednej) form polskich. Jest to typowy przykład ilustrujący charakterystyczne dla języka polskiego niedopowiedzenie językowe. W przykładzie (20'') podajemy dosłowne tłumaczenie zdania litewskiego, w którym wyraźnie uwidocznione zostają nieujawnione na płaszczyźnie formalnej języka polskiego treści. Gdybyśmy nawet przyjęli, że w tym konkretnym przypadku pojedyncze użycie wykładnika modalności imperceptywnej w zdaniu polskim

może funkcjonalnie odpowiadać znaczeniu wnoszonym przez sześć wykładników w języku litewskim, to i tak powstaje pytanie, dlaczego marker imperceptywności w polskim zdaniu pojawia się dopiero przy piątym czasowniku *mówić*. Można również zapytać, jak w języku polskim wyznaczane są granice fragmentów nacechowanych i nienacechowanych imperceptywnie. Analiza zasobów polskojęzycznych, wzbogaconych o rzutowane na podstawie zasobów litewskich granice części nacechowanych imperceptywnie, może prowadzić do ujawnienia niezaważanych dotychczas cech języka polskiego.

Podsumowanie

Zespół lingwistów Instytutu Slawistyki PAN od lat tworzy i publikuje na stronach CLARIN-PL anotowane korpusy języków bałtyckich i słowiańskich. Przedstawicielami tych korpusów są Polish-Lithuanian Parallel Corpus „2” oraz Polish-Russian Parallel Corpus, które w tym artykule zostały zastosowane do analizy i ilustracji poszczególnych zagadnień głównie w polsko-litewskiej konfrontacji.

O języku litewskim można powiedzieć, że ma strukturę bardziej zachowawczą, która od stuleci podlega pewnym przekształceniom, lecz zdecydowanie mniejszym niż ma to miejsce w języku polskim. Czytelniejsza struktura morfologiczna form litewskich sprzyja regularności użyciu konkretnych wyrażen, form, morfemów, czym gwarantowana jest mniejsza niż w języku polskim liczba niedopowiedzeń językowych. Tę własność języka litewskiego można zastosować w badaniach korpusowych poprzez rzutowanie jednoznacznych w języku litewskim znaczeń na skorelowane z nimi w korpusie teksty polskojęzyczne w celu selekcji polskich zasobów do przeprowadzenia konkretnych badań.

Bibliografia

- Ambrasas, V. (1990). *Sravnitel'nyi sintaksis prichastiŭ baltiŭskikh iazykov*. Vil'nius: Mokslas.
- Anstatt, T. (2008). Der slavische Vokativ im europäischen Kontext. W: L. Geist, G. Mehlhorn (red.), *Linguistische Beiträge zur Slavistik XV*, s. 9–26, München: München Sagner.
- Čermák, F., Rosen, A. (2012). *The case of InterCorp, a multilingual parallel corpus*. *The International Journal of Corpus Linguistics*, 17, s. 411–427.
- Duszkina, M., Roszko, D., Roszko, R. (2021). New Parallel Corpora of Baltic and Slavic Languages — Assumptions of Corpus Construction. *Lecture Notes in Artificial Intelligence. LNAI 12848*, s. 172–183, DOI: 10.1007/978-3-030-83527-9.
- Grzegorzczak, R. (1979). *Zarys słownictwa polskiego: Słownictwo opisowe*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Koseska-Toszecka, V., Korytkowska, M., Roszko, R. (2009). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie Dialog.
- Krongauz, M. A. (1999). Obrashcheniya kak sposob modelirovaniia kommunikativnogo prostranstva. W: *Logicheskiŭ analiz iazyka: Obraz cheloveka v kul'ture i iazyke* (s. 130–131). Moskva: Indrik.

- Linguee*. Portal tłumaczeniowy Linguee: <https://www.linguee.com/>.
- Polonskii, A. V. (2002). Egotiv, vokativ, nominativ: sub"ekt i padezhnaia paradigma. *Russkii iazyk za rubezhom*, 3, s. 27–35. Pobrane z: http://gramota.ru/biblio/magazines/ryzr/rzr2001-03/28_197 (dostęp: 1.09.2020).
- Przepiórkowski, A., Bańko, M., Górski, R., Lewandowska-Tomaszczyk, B. (red.). (2012). *Narodowy korpus języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (na tle literackich języków polskiego i litewskiego)*. Warszawa: Instytut Slawistyki PAN.
- Roszko, D., Roszko, R. (2021). Korpusy wielojęzyczne wkładem Instytutu Slawistyki PAN w rozwój infrastruktury CLARIN-PL. Przykłady analizy korpusowej nad wołaczem. W: J. L. Banasiak, A. Kikiewicz, J. Mazurkiewicz-Sułkowska (red.), *Języki słowiańskie dziś — w kręgu kategorii, struktur i procesów* (s. 281–313). Warszawa–Łódź: Instytut Slawistyki PAN – Wydawnictwo Uniwersytetu Łódzkiego.
- Roszko, R., Roszko, D. (2018). *Polish-Lithuanian Parallel Corpus „2”*, CLARIN-PL digital repository. Pobrane z: <http://hdl.handle.net/11321/539> (dostęp: 1.09.2020).
- Smoczyński, W. (2007). *Słownik etymologiczny języka litewskiego: Lietuvių kalbos etimologinis žodynas*. Wilno: Wydawnictwo Uniwersytetu Wileńskiego.
- Smoczyński, W. (2019). *Słownik etymologiczny języka litewskiego. Wydanie drugie, poprawione i znacznie rozszerzone*. Pobrane z: <https://www.rromanes.org/pub/alii/Smoczy%20W.%20S%20ownik%20etymologiczny%20j%20c%99zyka%20litewskiego.pdf> (dostęp: 1.09.2020).
- Suprun, V. I. (2001). Antroponimy v vokativnom upotreblenii. *Izvestiia Ural'skogo gosudarstvennogo universiteta*, 20, s. 92–96. Pobrane z: http://www.philology.ru/linguistics2/suprun_v-01.htm (dostęp: 1.09.2020).