Danuta Roszko, Roman Roszko

**O zastosowaniach *Eksperymentalnego korpusu
polsko-litewskiego* w badaniach nad polszczyzną
współczesną oraz w teoretycznych polsko-
-litewskich badaniach konfrontatywnych**

0. Wstęp

Korpusy dwu- i wielojęzyczne pojawiły się jako wartość dodana do korpusów jednojęzycznych. Eksperymentalny dwujęzyczny korpus polsko-litewski (EKorpPL-LT) można teoretycznie traktować jak dwa korpusy jednojęzyczne (polski i litewski), jak również korzystać z każdego z nich niezależnie od drugiego. Jednak ustanowione ściśle wzajemne powiązania między polską i litewską bazą pozwalają widzieć EKorpPL-LT jako narzędzie bardziej zaawansowane niż tylko dwa jednojęzyczne korpusy. Przeprowadzone przez autorów eksperymenty badawcze potwierdzają, że dwu- i wielojęzyczne korpusy to nie tylko baza leksykalna dla tłumaczy, materiał empiryczny służący ekstrakcji złożonych danych lingwistycznych, ustalaniu algorytmów, miary bliskości znaczeniowej słów (z jednego języka oraz pomiędzy polskim i litewskim), definiowaniu metod automatyzujących przekład, lecz również doskonałe źródło wiedzy o każdym języku z perspektywy innego.

W artykule zostają przedstawione wyniki kilku eksperymentów, które potwierdzają tezę, że istnieją takie fakty językowe, których konstatacja jest utrudniona bez odniesienia do innego systemu o analogicznej w swych podstawach strukturze. Uwaga autorów zostaje skierowana na użycie wołacza oraz zaimków dualnych, na jakość przekładów instrukcji obsługi, a także na wybrane zagadnienia związane z kwantyfikacją zakresową.

1. Eksperymentalny dwujęzyczny korpus polsko-litewski¹

W internecie znajdują się bogate zasoby organizujące teksty polskojęzyczne (np. Narodowy Korpus Języka Polskiego – <http://nkjp.pl/>; Korpus Instytutu Podstaw Informatyki PAN – <http://korpus.pl/>, Korpus Języka Polskiego Wydawnictwa Naukowego PWN – <http://korpus.pwn.pl/>) oraz teksty litewskojęzyczne (np. Korpus współczesnego języka litewskiego „Tekstynas”² – <http://tekstynas.vdu.lt/>). Niestety nie ma korpusów, które tworzyłyby wartościową bazę tekstów połączonych (polskich i litewskich). Fakt, powstały wielojęzyczne korpusy, które włączają zasoby obu języków do wspólnej bazy, np. EMEA (<http://opus.lingfil.uu.se/EMEA.php>), KDE4 (<http://opus.lingfil.uu.se/KDE4.php>), OPUS (<http://opus.lingfil.uu.se/>), ParaSol corpus (<http://parasol.unibe.ch/>). Jednak objętość tych korpusów oraz charakter ich zasobów (np. teksty specjalistyczne medyczne oraz ustawodawstwo unijne) nie pozwalają na prowadzenie zaawansowanych prac, takich jak badania konfrontatywne czy budowa wielojęzycznych słowników.

1.2. Struktura korpusu. EKorpPL-LT powstaje³ z myślą o (a) budowie „pamięci tłumacza” do zastosowań w pracy translatorskiej⁴, (b) polsko-litewskich badaniach konfrontatywnych, oraz (c) planowanej budowie słownika polsko-litewskiego.

1.2.1. Gromadzone przez lata w formie elektronicznej podstawy tłumaczeń oraz ich przekłady posłużyły za podstawę pierwszego pionu korpusu, zwanego umownie podkorpusem A. Wszystkie zawarte w tej części zasoby zostały zrównoleglone z wykorzystaniem narzędzi z licencją otwartą TextAlign (by An-

¹ EKorpPL-LT – to dwa niezależne korpusy: równoległy i porównawczy. W tym artykule każde odniesienie do EKorpPL-LT oznacza tylko korpus równoległy (paralelny).

² Nazwa w oryginale: Dabartinės lietuvių kalbos tekstynas.

³ Korpus znajduje się w ciągłej rozbudowie.

⁴ Autorzy od blisko 20 lat zajmują się przekładem.

drew Manson)⁵ oraz programów komercyjnych Nova Text Aligner i Terminotix AlignFactory. Po zrównolegleniu zasoby zostały załadowane do programu Athel ParaConc (<http://www.athel.com/para.html>) oraz dodatkowo stworzono bazy terminów z wykorzystaniem ApSIC Xbench (http://www.apsic.com/en/products_xbench.html) oraz Terminotix LogiTerm Pro (http://www.terminotix.com/index.asp?name=LogiTerm_Pro&content=item&brand=2&item=12&lang=en).

W podkorpusie A znalazły się dokumenty jak: Kodeks Cywilny Republiki Litewskiej, poszczególne ustawy (np. o podatku dochodowym od osób fizycznych), rozporządzenia, projekty europejskie (partnerskie), wykazy leków, działalności, wypisy skrócone i rozszerzone z rejestru osób prawnych, typowe formularze stosowane w urzędach celnych, skarbowych, cywilnych, w policji, w zakładach ubezpieczeń społecznych, w firmach ubezpieczeniowych. Ponadto – dokumentacja medyczna (np. epikryzy), sądowa (np. wnioski, postanowienia, wyroki, korespondencja itd.), biznesowa i techniczna (np. umowy bilateralne, warunki przetargu, pisma procesowe, certyfikaty, opisy techniczne, instrukcje eksploatacji i obsługi, gwarancje, wymogi techniczne, regulaminy, plany, foldery, dokumentacje zdawczo-odbiorcze) itd. Istotną cechą podkorpusu A jest fakt, że ograniczono liczbę sztamowych dokumentów (takich jak dowód rejestracyjny pojazdu, międzynarodowy list przewozowy CMR, dowód osobisty, paszport, akt urodzenia, akt zgonu, niektóre umowy i inne dokumenty o narzuconej ustawodawstwem unijnym formie) do maksymalnie dwóch egzemplarzy.

Zasoby podkorpusu A nie zostały poddane lematyzacji i anotacji. Jego objętość wynosi około sześciu milionów tokenów.

1.2.2. W roku 2010 przystąpiono do tworzenia drugiego pionu EKorpPL-LT w oparciu o przekłady beletrystyki i inne teksty ogólnie dostępne w internecie. Tak powstał podkorpus B⁶, liczący blisko dwa miliony tokenów. Podkorpus B został opracowany z myślą o prowadzonych badaniach własnych nad językami polskim i litewskim (badania konfrontatywne) oraz w związku z rozszerzeniem bazy terminologicznej, leksykalnej na cele planowanego polsko-litewskiego słownika.

W podkorpusie B zwrócono uwagę na właściwe zrównoważenie tekstów. Dlatego do zasobów zostały włączone różnorodne materiały, reprezentujące możliwie szeroki zakres tematyczny. Poza tekstami literackimi, technicznymi, medycznymi, prawniczymi, sądowniczymi znalazły się materiały związane

⁵ Program był używany tylko na początkowym etapie prac. Ze względu na jego ograniczone możliwości zdecydowano się na wykorzystanie programów komercyjnych.

⁶ Podkorpus A zawiera teksty oryginalne i ich tłumaczenia wykonane przez autorów artykułu. Podkorpus B jest wolny od tego typu tłumaczeń.

z nowymi technologiami i osiągnięciami cywilizacyjnymi. Kierując się zasadą wewnętrznego zrównoważenia zasobów, odstąpiono od niekontrolowanego włączenia powszechnie dostępnych zasobów ustawodawstwa unijnego. Ograniczono się do kilku ważnych z punktu widzenia zasobów leksykalnych aktów (szerzej na ten temat por. Roszko, Roszko 2013). Analogicznie zawężono liczbę tekstów medycznych (do kilku artykułów naukowych oraz opisowego atlasu anatomicznego), technicznych (zawierających instrukcje obsługi oraz foldery reklamowe), z branży turystycznej (jak przewodniki, prospekty) oraz różnego rodzaju poradników.

Osiemdziesiąt procent zasobów podkorpusu B stanowi beletrystyka. Starano się skompletować wzajemne polsko-litewskie tłumaczenia z okresu po II wojnie światowej, np. utwory A. Kuklysa, R. Černiauskasa, J. Šikšnelisa, E. Białołęckiej, S. Lema, W. Gombrowicza itd. Włączono również przekłady literatury światowej, np. powieści P. Coelho, J. K. Rowling itd. Niektóre utwory są reprezentowane w całości, niektóre w reprezentatywnej dla dzieła części. W chwili obecnej podkorpus B zawiera 47 utworów beletrystycznych w obu wersjach językowych.

Zasoby podkorpusu B zostały zrównoleglone. Część materiałów została poddana lematyzacji i anotacji z wykorzystaniem narzędzi, takich jak Morfeusz (<http://sgjp.pl/morfeusz/>) dla języka polskiego i Anotatorius (http://donelaitis.vdu.lt/main.php?id=4&nr=7_2) dla języka litewskiego. Wszystkie zasoby tego podkorpusu zostały załadowane do programu Athel ParaConc (<http://www.athel.com/para.html>).

EKorpPL-LT został szczegółowo opisany w Roszko, Roszko 2013. Również w tym tomie w artykule D. Roszko (s. ????) na przykładzie Eksperymentalnego korpusu gwary puńskiej w Polsce można prześledzić analogiczne do zastosowanych również w tym korpusie sposoby zapisu anotacji, lematyzacji oraz próbki tekstów zrównoleglonych.

2. Wołacz

Wprowadzenie do podkorpusu B anotacji morfosyntaktycznej pozwoliło zaobserwować, jak poszczególne klasy (podklasy) leksemów są rozpowszechnione w każdym z języków. Zaobserwowano wiele różnic, np. niską (w porównaniu z litewskim) frekwencję polskich form imiesłowowych, wysoką natomiast – form przyimkowych. Jednak nie każda dysproporcja w częstości użycia poszczególnych klas/podklas dla zestawianych języków o czymkolwiek przesądza. W wypadku polskiego wołacza można jednak wysnuć wniosek, że jest wypierany przez formę imienną w mianowniku, por.:

[1] *Iwan!* – cichutko zawołał Berlioz.

Ivanai! – negarsiai šuktelėjo Michailas Aleksandrovičius.

- [2] – *Marek!* – krzyknął prokurator i centurion wszedł pod kolumny.
 – *Markai!* – šūktelėjo prokuratorius, ir kenturionas įžengė po kolonomis.
- [3] – Halo, *Kelvin!* – odezwał się ochryple.
 – *Alio, Kelvinai!* – pašaukė kimiū balsu.

Oczywiście do takiego samego wniosku można dojść, analizując polskie teksty na przestrzeni ostatnich kilkudziesięciu lat. Jednak dopiero dane z korpusów równoległych uświadamiają badaczom, jak bardzo zaawansowane jest to zjawisko. Wystarczy porównać łączne liczby 719 polskich i 15 065 litewskich leksemów w wołaczach. Na przykład w powieści *Solaris* Stanisława Lema w wersji polskiej imię Kelvin pojawia się 48 razy. W tłumaczeniu na język litewski – 50 razy. W tabeli 1. podano rozkład użycia poszczególnych przypadków dla wspomnianego wyżej leksemu *Kelvin* w obu językach.

Tabela 1. Użycie poszczególnych form przypadków dla leksemu *Kelvin* w powieści Stanisława Lema *Solaris* oraz w tłumaczeniu na język litewski (wyróżniono tylko przypadki użyte w powieści).

Lp.	Przypadek	Liczba użyć w wersji	
		polskiej	litewskiej
1.	mianownik	46	7
2.	dopełniacz	2	2
3.	biernik	–	1
4.	wołacz*	–	40

* Nie wdajemy się w tym miejscu w dyskusję, czy wołacz jest jednym z przypadków, czy też osobną kategorią morfologiczną.

3. Zaimki dualne

W oparciu o zgromadzony w EKorpPL-LT materiał możliwe jest porównanie przekładów z języka polskiego na litewski, sporządzonych przez polskich i litewskich tłumaczy, pod kątem liczby użytych zaimków dualnych. Polscy tłumacze z założenia nie rozróżniają liczby osób równej dwóm od liczby większej od dwóch. Chyba że w polskim tekście znajduje ona formalne wyrażenie, np. *ich dwóch*, *we dwie*, *obaj* itd. Dla zilustrowania zjawiska również posłużymy się przykładami zaczerpniętymi z powieści *Solaris*:

- [4] – Gdzie *my* jesteśmy, Harey?
 – Kur *mudu* esame, Hare?

- [5] – Sartorius przekonał Snautą, że go oszukałeś. Teraz *oni* ciebie oszukają.
Pod pozorem montowania aparatury rentgenowskiej budują anihilator pola.
– Sartorius įtikino Snautą, kad *tu* apgavai jį. Dabar *jiedu* apgaus tave.
Dėdamiesi montuoja rentgeno aparatūrą, *jiedu* stato lauko anihiliatorių.
- [6] Znowu byłem w zupełnej ciemności, nie pojmowałem niczego, a przez chwilę zdawało mi się, że uda mi się go przekonać i wspólnymi siłami zaatakujemy zagadkę.
Aš vėl atsidūriau aklavietėje, nieko nesupratau. O jau buvau bemanąs, kad pasiseks jį įtikinti ir *mudu* bendrom jėgom pulsime šitą paslaptį.

W przykładzie [4] polskiemu *my* odpowiada litewski zaimek dualny *mudu*. Podobnie w [5] polskiemu *oni* odpowiada inny litewski zaimek dualny *jiedu*. W litewskich zdaniach z przykładów [5] i [6] występują kolejne formy zaimków dualnych *jiedu* i *mudu* (wyróżnione drukiem rozstrzelonym), które na płaszczyźnie formalnej nie mogą zostać powiązane z żadną polską formą, por.:

- [5a] $\emptyset$ budują anihilator pola
jiedu stato lauko anihiliatorių
- [6a] i $\emptyset$ wspólnymi siłami zaatakujemy $\emptyset$ zagadkę
ir *mudu* bendrom jėgom pulsime šitą paslaptį

Wstawienie form zaimkowych do tekstu litewskiego (por. [5a] i [6a]) nie jest warunkowane brakiem informacji o osobie i liczbie w formie litewskiego czasownika. Ta formalna dysproporcja między językami jest następstwem innego zjawiska zwanego niedopowiedzeniem kwantyfikacyjnym (por. Koseśka, Gargov 1990). Niedopowiedzenie kwantyfikacyjne – typowe dla języka polskiego – oznacza, że aktualne znaczenie kwantyfikacyjne ujawnia się tylko w kontekście i/lub w sytuacji. Jak można zaobserwować, w polskim wariancie przykładu [6a] dwukrotnie pojawia się $\emptyset$ jako ekwiwalent litewskich zaimków osobowego – *mudu* i wskazującego – *šis* ‘ten’. W wyżej podanej pełnej wersji zdania [6] jest obserwowana analogiczna sytuacja, w której litewskiemu zaimkowi osobowemu *aš* ‘ja’ odpowiada na płaszczyźnie formalnej polskie $\emptyset$. Zjawisko kwantyfikacji w obu językach jest przedmiotem opisu w kolejnym punkcie.

4. Określoność–nieokreśloność

4.1. Określoność. W literaturze przedmiotu wyrażanie określoności w języku litewskim łączy się z postpozycyjnym zaimkiem *-jis*, *-ji*, por. gramatyki języka litewskiego (Ambrasas 1997; Valeckienė 1998). Formy zawierające morfem *-jis*, *-ji* uznaje się za człon nacechowany opozycji, por. *garsusis* ‘ten

głośny’ wobec *garsus* ‘(jakiś) głośny’. Pary analogiczne do podanej tworzą morfologiczną kategorię określoności⁷. Litewska kategoria określoności, choć podobna w swej strukturze do postaci znanej chociażby z języka bułgarskiego, ma pewne ograniczenia. Formy określone mogą być utworzone tylko dla przymiotników (w głównej mierze jakościowych), liczebników porządkowych, zaimków osobowych, wskazujących, dzierżawczych oraz imiesłowów odmiennych czynnych i biernych. Wśród wymienionych klas nie ma rzeczowników. Te bowiem nie łączą się z postpozycyjnym *-jis*, *-ji*. Ograniczenie w łączliwości do form imiennych różnych od rzeczownika może oznaczać zawężoną funkcjonalność tej kategorii, w konsekwencji – prowadzić do niedopowiedzenia kwantyfikacji w przypadku użyć wyrażen prostych fundowanych przez rzeczowniki. Dopiero analiza zasobów EKorpPL-LT wykazała, że to formalne ograniczenie bezpośrednio nie przekłada się na wzrost wieloznaczności kwantyfikacyjnej. Spodziewane w języku litewskim wyrażenia proste rzeczownikowe są zastępowane przez inne imiesłowowe (por. [7], [8]) lub wyrażenia złożone fundowane przez imiesłowy (por. [9]) lub rzeczowniki. Te ostatnie koniecznie muszą zawierać zaimek w prepozycji (por. [10]).

[7] [...] w tym miejscu *mówiący* pozwolił sobie na uśmiech.

[...] ir čia *kalbantysis* išdrįso nusišypsoti.

[8] Przybysz szepnął coś *siedzącemu* [...]

Atvykėlis kažką pakuždėjo *sėdinčiajam* [...]

[9] *Dane, które zebrał*, potwierdziły ostatecznie [...]

Surinktieji duomenys galutinai patvirtino [...]

[10] *Kobieta z zieloną ceratową torbą* wyszła z klatki schodowej na podwórko.

Toji moteris, nešina žaliu ceratiniu krepšeliu, išėjo pro laukujės duris į kiemą.

Analiza EKorpPL-LT ujawnia również polsko-litewskie odpowiedniości typu polska nazwa własna – litewski zaimek wskazujący (por. też niżej przykład [21]):

[11] Ale *Andrzej* trzęsący się ze strachu w ogóle nie jest podobny do ciebie.

Tačiau *šitas* tirtantis iš baimės visai nepanašus į tave.

[12] – Dzień dobry – powiedział *Mały Książę*.

– Laba diena, – tarė jam *šis*.

Liczba użyć zaimków w języku litewskim jest większa niż w języku polskim, por. przykłady [10]–[12] oraz wcześniejsze, [5], [5a], [6], [6a]. Wysoka częstość użycia form z postpozycyjnym *-jis*, *-ji* oraz zaimków w postaci wyrażen prostych

⁷ Tak przyjęto nazywać tę kategorię w litewskiej literaturze przedmiotu.

lub wchodzących w skład wyrażen złożonych jest przejawem konsekwentnego rozgraniczania znaczeń kwantyfikacji na poziomie formalnym. Język polski – jak już to było wspomniane wyżej – w większym stopniu cechuje niedopowiedzenie kwantyfikacyjne.

Cechą form z postpozycyjnym *-jis*, *-ji* jest również wyrażanie znaczeń ogólnych zwyczajowych (por. termin w: Roszko 2004 oraz przykłady niżej w p. 4.2.3).

Badania prowadzone z wykorzystaniem EKorpPL-LT potwierdziły wysoką częstotliwość użycia litewskiego zaimka wskazującego *tas*, *ta* ‘dokładnie ten’, który najczęściej pojawia się tam, gdzie użycie postpozycyjnego *-jis*, *-ji* nie jest możliwe. Litewskiemu *tas*, *ta* w 70% odpowiada ekwiwalentne użycie polskich *ten*, *to*, *tamten*, *ów*.

- [13] Wszyscy prócz nieruchomego procuratora odprowadzali spojrzieniami Marka Szczurzą Śmierć, który skinął *na aresztanta*, każąc *mu* iść za sobą. Ir visi, išskyrus nejudantį prokuratorių, palydėjo žvilgsniu Marką Žiurkiąmušį, kuris mostelėjo ranka *suimtajam*, rodydamas, kad *tas* turi sekti jam įkandin.
- [14] Alojzy Mogarycz działa do dziś *na stanowisku*, którego zrzekł się Rimski, to znaczy na etacie dyrektora finansowego Varietes. Šis ne tik buvo, bet ir dabar dar gyvas, ir eina kaip tik *tas* *pareigas*, kurių atsisakė Rimskis, jis – Varjetė finansų direktorius.
- [15] A jak on się nazywa? – zapytano go cicho, na ucho.
O kokia jo pavardė? – neatlyžo *tas* pats tylus balsas.

Analogiczne funkcje do opisanego lit. *tas*, *ta* posiadają również inne litewskie zaimki wskazujące: *šis*, *ši* oraz *šitas*, *šita*. Przy czym formy *šis*, *ši* wydają się tylko częściowym odpowiednikiem polskich *ten*, *tamten*, *ów* (w oparciu o EKorpuPL-LT ta zgodność wynosi 40%), np.:

- [16] Obywatele, *ten* *chory* zostanie przyjęty poza kolejką.
Piliečiai, *šis* *ligonis* užeis be eilės.
- [17] Wybaczyć nie mogę – stanowczo odparł *tamten*.
Negaliu dovanoti, – griežtai tarė *šis*.
- [18] *Ów* skłonił się, kiedy go odbierał, i schował sakiewkę pod płaszczem.
Šis nusilenkė, priimdamas maišelį, ir pasikišo jį po apsiaustu.

Pozostałym użyciom *šis*, *ši* odpowiada zazwyczaj $\emptyset$ (por. [19], [20]), sporadycznie inny sposób wyrażenia określoności (por. nazwę własną w przykładzie [21]) lub wyrażenie z niedopowiedzeniem kwantyfikacyjnym (por. [22]).

- [19] Fosforescencja oceanu była zjawiskiem już opisanym.
*Šis reiškiny*s, okeano fosforescencija, buvo jau aprašytas.

- [20] Czy wojna między kwiatami a barankami nie jest rzeczą poważną?
Argi tai ne svarbu, kad vyksta *šis* karas tarp avių ir gėlių?
- [21] **Arkadiusz Apołłonowicz** został bowiem momentalnie przeniesiony do Briańska na stanowisko kierownika punktu skupu runa leśnego.
Šis mikliai buvo permestas į Brianską ir paskirtas grybų paruošų punkto vedėju.
- [22] – Nie pojedę! – krótko oświadczył *kierowca*.
– Nevešiu! – trumpai tarę *šis*.

Niemal wszystkim użyciom litewskich *šitas*, *šita* odpowiada polski zaimek *ten*⁸, por.:

- [23] *To pudelko* jest ze szczerego złota, widać po wadze.
Šitas daiktelis gryno aukso, iš svorio jaučiu.

4.2. Nieokreśloność

Na nieokreśloność składają się dwie podklasy – egzystencjalność i ogólność (por. Koseska, Gargov 1990). Każda z podklas podlega uszczegółowieniu. Egzystencjalność występuje w dwóch wariantach (a) presuponującej jednostkowość⁹ (w terminologii stosowanej przez Karolaka (2001) odpowiada to użyciu eluzywnemu), (b) właściwej (odpowiednio u Karolaka – użycie tetyczne). Również ogólność ma dwa warianty – zwyczajowy (u Karolaka – użycie hipotetyczne) oraz właściwy (u Karolaka – użycie antytetyczne).

W poniższych punktach zostanie zwrócona uwaga na wybrane szeregi zaimków w polskim i litewskim, które przynajmniej w jednym języku można uznać za wyspecjalizowany wykładnik egzystencjalności lub ogólności.

4.2.1. Egzystencjalność presuponująca jednostkowość

W obu językach można wyróżnić wykładniki, które pretendują do wyrażania znaczeń egzystencjalnych presuponujących jednostkowość. Są to: polski *pewien* i litewski *tam tikras*. Dla ponad 70% procent wystąpień litewskiego *tam tikras* polskim odpowiednikiem jest wyrażenie złożone zawierające pol. *pewien*, np.:

- [24] Potrzeba mu do doświadczeń *pewnej* ilości *plazmy*, jako wyjściowego budulca F.

Tyrimams jam reikia *tam tikro kiekio plazmos*, kaip pirminės F medžiagos.

W pozostałych wypadkach (około 30%) pojawiają się polskie leksemy *określony*, *jakiś*, np.:

⁸ Użyty tylko w wyrażeniach złożonych, np. *to serce* lub zredukowanych, np. *to (serce)*.

⁹ Na temat wariantów egzystencjalności i ogólności, por. Roszko 2004.

- [25] Gdy problem udało się, w *jakiejs* przynajmniej *mierze*, rozwikłać, okazało się, że [...]
 Kai šią mįslę pasisekė bent *tam tikru mastu* įminti, paaiškėjo, kaip [...]

Jednak gdy na ten sam problem spojrzymy przez pryzmat polskiego *pewien*, to okaże się, że w przeważającej większości jego ekwiwalentem nie jest litewska forma *tam tikras*, lecz *vienas* ‘jakiś’ [dosłownie: ‘jeden’], np.:

- [26] Przedwczoraj wieczorem w pobliżu świątyni poznałem *pewnego* młodego człowieka, który przedstawił mi się jako Juda z Kiriatu.
 Buvo taip, – noriai ėmė pasakoti suimtasis, – užvakar vakare prie šventyklos susipažinau su *vienu jaunuoliu*, pasivadinusiu Juda iš Kiriato miesto.
- [27] Mam, co prawda, *pewną* teorię na ten temat.
 Turiu, tiesa, *vieną* teoriją tuo klausimu.
- [28] [...] lecz wkrótce Vincukas dostał cios w serce: gdy *pewnego* ranka szykował się do wyjścia z domu, żona ostrzegła go, aby nie jechał na działkę, [...] [...] tačiau netrukus Vincukas gavo smūgį į paširdžius: kai *vieną* rytą ruošėsi išeiti iš namų, žmona perspėjo, kad į sodą nevažiuotų, [...]

Sporadycznie notuje się zgodność użycia pol. *jeden* i lit. *vienas* w znaczeniu egzystencjalnym (właściwym), por.:

- [29] [...] a raz *jeden* otarł się jedwabistą z wyglądu powierzchnią o szkło, [...] [...] o kartą *vienas* slystelėjo per stiklą savo šilkiniiu paviršiumi. [...]

Zaobserwowane dysproporcje między użyciami poszczególnych polsko-litewskich ekwiwalentów: litewskiemu *tam tikras* odpowiadają pol. *pewien*, *określony* (sporadycznie *jakiś*), polskiemu zaś *pewien* – lit. *vienas* (przede wszystkim) i *tam tikras*, pozwalają podejrzewać, że użyciu polskiego leksemu *pewien* towarzyszy niedopowiedzenie kwantyfikacyjne. Oznacza to, że w zależności od kontekstu pol. *pewien* może wyrażać (a) znaczenia egzystencjalności presuponującej jednostkowość (wówczas litewskim ekwiwalentem jest *tam tikras*) lub (b) znaczenia egzystencjalności właściwej (wówczas koresponduje z litewskim wykładnikiem *vienas*).

4.2.2. Egzystencjalność właściwa

Polsko-litewska para ekwiwalentów *pewien* – *vienas* została po części opisana w p. 4.2.1. W tym miejscu należy zauważyć, że litewskie *vienas* jest wykładnikiem nie wnoszącym do przekazu niedopowiedzenia kwantyfikacyjnego. Polskie *pewien* jest jednak pod tym względem dwuznaczne. Właściwa interpretacja wyrażenia zawierającego pol. *pewien* wymaga kontekstu lub sytuacji, niekiedy minimalnej wspólnej dla nadawcy i odbiorcy wiedzy o świecie, jak w poniższym przykładzie:

[30] Czy pomyślałeś, że *pewnego* dnia ciebie już nie będzie?

Ar pagalvojai, kad *vieną* dieną tavęs nebebus?

Analiza EKorpPL-LT pod kątem regularnych wykładników egzystencjalności właściwej ujawnia litewską serię zaimków z częstką *kaž-* (*kaži-*, *kažin*), której z reguły odpowiada polska seria zaimków z częstką *-ś*, np.:

[31] Pewno mu się zdawało, że *coś* widzi... – mówiłem coraz wolniej, patrząc mu w oczy.

Jam turbūt atrodė, kad *kažką* mato... – kalbėjau aš vis lėčiau, žvelgdamas jam į akis.

[32] Na balkonie *jakiś* *wstrząśnięty* obywatel znalazł paczkę banknotów w bankowej banderoli z napisem: „Jeden tysiąc rubli”.

Balkone *kažkoks* *sutrikęs pilietis* rado savo kišenėje perjuostą, lyg tiesiai iš banko, pluoštelį su užrašu: „Vienas tūkstantis rublių”.

W ponad 40% użyć litewskiej serii zaimków z częstką *kaž-* (*kaži-*, *kažin*) nie odpowiada żaden polski zaimek. W polskiej wersji rejestruje się wyrażenie proste lub złożone z niedopowiedzeniem kwantyfikacyjnym, np.:

[33] I zobaczyłem *niezwykłego* *małego* *człowieka*, który bacznie mi się przyglądał.

Ir pamačiau, kad į mane oriai žiūri *kažkoks* *nepaprastas berniukštis*.

[34] Lecz on rzucił *czar* na dom.

Bet jis gaubė tą visą namą *kažkokiais* *burtais*.

Aby sprawdzić, czy polskie zaimki z częstką *-ś* są funkcjonalnymi ekwiwalentami litewskich zawierających częstkę *kaž-*, należy sprawdzić, jakie litewskie formy odpowiadają użyciom polskiej serii zaimków z częstką *-ś* (*ktoś*, *coś*, *jakiś*, *któryś* itd.). Okazało się, że w 50% obserwowana jest zgodność: polskiemu wyrażeniu prostemu lub złożonemu zawierającemu zaimek z częstką *-ś* odpowiada litewskie wyrażenie z częstką *kaž-*. W pozostałych wypadkach są to litewskie zaimki pytajno-względne, np.:

[35] – Macie przecież *jakiś* *plan* *badania*?

– Juk turite *kokį* *tyrinėjimų planą*?

albo seria zaimków zawierająca częstkę *nors*, np.:

[36] Znasz *jakiś* *język* prócz aramejskiego?

O dar *kokią* *nors* *kalbą*, be aramėjų?

Podobnie jak w punkcie 4.2.1. również i w tym wypadku zauważana jest dysproporcja między ekwiwalentami. Litewskiej serii zaimków z częstką *kaž-* (*kaži-*, *kažin*) odpowiada polska seria zaimków z częstką *-ś*. Natomiast polskiej serii zaimków z częstką *-ś* mogą odpowiadać litewskie zaimki z częstką *kaž-*

(*kaži-*, *kažin*), z częstką *nors* lub zaimki pytajno-względne w czystej postaci (bez dodatkowych części). Z powyższego zestawienia wypływa wniosek, że polska seria zaimków z częstką *-ś* dopuszcza niedopowiedzenie kwantyfikacyjne. W pewnych kontekstach i sytuacjach polskie wyrażenia proste i złożone zawierające tę właśnie serię polskich zaimków (np. *ktoś*, *jakiś* itd.) mogą być wykładnikiem kwantyfikacyjnych znaczeń egzystencjalności właściwej, w innych zaś – wykładnikiem ogólności zwyczajowej.

4.2.3. Ogólność zwyczajowa

Pierwszym wyróżniającym się wykładnikiem tych znaczeń są opisywane w punkcie 4.1. litewskie formy z postpozycyjnym *-jis*, *-ji*, np.:

[37] *Pierwszy nauczyciel* zapada głęboko w sercu.

Pirmasis mokytojas visada išlieka širdyje.

[38] *Prawdziwy kierowca* zawsze siada za kierownicą i, jeśli nie ma centralnego zamka, ręką od środka otwiera pasażerowi drzwi.

Tikrasis vairuotojas sėda prie vairo ir, jei nėra centrinio užrakto, ranka iš vidaus atidaro kitas dureles.

[39] *Cicha woda* brzegi rwie.

Tylioji kiaulė giliai šaknį knisą.

Polskich odpowiedników litewskich form zawierających postpozycyjne *-jis*, *-ji* nie wyróżnia żaden element formalny. Tym samym polskie wyrażenia, z typowym dla siebie niedopowiedzeniem kwantyfikacyjnym, wymagają ścisłej kooperacji z kontekstem celem ustalenia aktualnego znaczenia kwantyfikacyjnego.

Drugi regularny wykładnik ogólności zwyczajowej w języku litewskim to formy zaimkowe zawierające częstkę *nors*, np.:

[40] – Musiałeś go *gdzieś* zostawić – rzekł Kubuś Puchatek.

– *Ktoś* musiał mi go zabrać – powiedział Kłapouchy. – I jak tu mieć dla nich serce? – dodał po dłuższej chwili milczenia.

– Būsi ją *kur nors* palikęs, tarė Pūkuotukas.

– *Kas nors* bus pasiėmęs, – pasakė Nulėpausis. – Va kokie, – pridūrė ilgokai patylėjęs.

[41] – Kiedy się *czegoś* pragnie, wtedy cały wszechświat sprzysięga się, byśmy mogli spełnić nasze marzenie – powiedział Alchemik, powtarzając słowa Starego Króla.

– Kai tikrai *ko nors* trokšti, visas pasaulis slapta padeda tau įgyvendinti tą troškimą, – senojio karaliaus žodžiais atsako Alchemikas.

Z przedstawionych powyżej polsko-litewskich par zdań wynika, że polskim odpowiednikiem dla litewskich wyrażen z częstką *nors* są wyrażenia z post-

pozycyjnym -ś (np. *ktoś, coś, jakoś, któryś* itd.). W punkcie 4.2.2. wskazano jednak, że te polskie formy mogą również korelować z litewskimi wykładnikami zawierającymi częstkę *kaž-* (*kaži-, kažin*). To już kolejny przypadek, kiedy polski wykładnik wymaga ścisłej kooperacji z kontekstem i sytuacją. Inaczej poza kontekstem nie jest się w stanie stwierdzić, czy konkretne polskie wyrażenia z postpozycyjnym -ś zostają skwantyfikowane egzystencjalnie czy też ogólnie.

W oparciu o EKorpPL-LT ustalono, że obok odpowiedników lit. *nors* – pol. -ś, możliwa jest również inna para: lit. *nors* – pol. -*kolwiek* (dla blisko 20% użycia litewskiego *nors*), por.:

- [42] Tu, żeby zobaczyć *cokolwiek* naprawdę, trzeba by odbiec, cofnąć się w jakąś olbrzymią dal, [...]
Čia, norint *ką nors* dorai pamatyti, reikėtų nubėgti, atsitraukti į begalinį tolių, [...]

Jednak gdy przyjmie się odwrotny kierunek poszukiwania w korpusie, to okaże się, że pol. formy z częstką -*kolwiek* są również odpowiednikiem litewskich wyrażen zawierających częstkę *bet*, por.:

- [43] – A gdzież on ma iść?
– *Gdziekolwiek*, prosto przed siebie.
– O kur, manai, ji gali nueiti!
– *Bet kur*. Nosies tiesumu...

O tej odpowiedniości jednak poniżej.

4.2.4. Ogólność właściwa

Przedstawiona w przykładzie [43] polsko-litewska odpowiedniość obrazuje typowy dla obu języków sposób oddania znaczeń kwantyfikacji ogólnej właściwej. Przy czym należy podkreślić, że polskie formy z -*kolwiek*, w różnych kontekstach mogą wyrażać dwa rodzaje ogólności. Ta własność nie jest typowa dla litewskiego odpowiednika *bet*. Tym samym po raz kolejny okazuje się, że włączone tu do opisu polskie wykładniki określoności-nieokreśloności cechuje niedopowiedzenie kwantyfikacyjne. Natomiast ich litewskie odpowiedniki charakteryzuje wysoki stopień specjalizacji, w następstwie czego określenie rodzaju kwantyfikacji dla danego litewskiego wyrażenia jest możliwe poza kontekstem i sytuacją.

4.3. W szerszym kontekście o określoności-nieokreśloności w polskim i litewskim

Przedstawione w punktach 4.1–4.2.4 polsko-litewskie ekwiwalenty wykładników określoności-nieokreśloności wskazują na wieloznaczność

polskich form i jednoznaczność litewskich. Na przykład polskie *pevien* w zależności od kontekstu może być wykładnikiem egzystencjalności presuponującej jednostkowość lub egzystencjalności właściwej. Z kolei polskie formy z charakterystyczną dla siebie częstką *-ś* są wykładnikiem egzystencjalności właściwej albo ogólności zwyczajowej. Podobnie polskie wykładniki zakończone na *-kolwiek* to nie tylko środki służące wyrażeniu ogólności zwyczajowej, lecz również ogólności właściwej. Litewskie odpowiedniki charakteryzuje użycie zawężone do jednego tylko zakresu znaczeń kwantyfikacyjnych.

Powyższy opis mógłby sugerować, że wszystkie polskie wykładniki określoności-nieokreśloności są wieloznaczne. Tak jednak nie jest. Aby potwierdzić tę tezę, wystarczy podać, że polskie nazwy własne, zaimki osobowe są jednoznacznym wykładnikiem określoności, natomiast leksemy jak: *każdy, wszyscy, wszędzie, zawsze, nigdy, nikt* itd. – jednoznacznym wykładnikiem ogólności właściwej, a: *nie każdy, nie wszyscy, nie wszędzie* itd. – wykładnikiem egzystencjalności właściwej. Oczywiście analogiczne zjawisko jest obserwowane w języku litewskim.

Rola EKorpPL-LT na tym się nie kończy. Autorzy widzą możliwość szerszego zastosowania faktu wąskiej specjalizacji litewskich wykładników określoności-nieokreśloności w ustalaniu zalecanych parametrów anotacji semantycznej w korpusach, na dalszym etapie – w opracowywaniu algorytmów na cele maszynowego przekładu.

5. Język instrukcji

Niska jakość polskich tłumaczeń różnego typu instrukcji dostarczanych wraz z towarem jest powszechnie znana. Panuje opinia, że tłumaczenia są zlecane osobom słabo znającym język polski lub wręcz otrzymywane w wyniku maszynowego przekładu w oparciu o niskiej jakości korpusy. Może to oznaczać cięcie kosztów przez producentów, importerów. Ale czy na pewno tylko to? Odpowiedź na to pytanie znajduje się pod koniec tego punktu.

W EKorpPL-LT wydzielono dział z tekstami różnych instrukcji, ulotek dołączanych do produktów powszechnego użytku (popularnie znanych pod postacią skrótców AGD i RTV). Analiza tych zasobów wykazała, że litewskie wersje instrukcji są lepiej przetłumaczone. Oto kilka przykładów pochodzących z instrukcji tego samego produktu:

[44] Dziękujemy *Państwu* za wybór [...]

Dėkojame, kad *pasirinkote* [...]

- [45] Uruchamiając [...] po raz pierwszy, **proszę wykonać** następujące czynności:
Pirmąkart paleidę [...], **turite** atlikti toliau aprašomus veiksmus.
- [46] **Możesz** usunąć tylko grupy POI, które sam **utworzyłeś**.
Galite šalinti tik pačių sukurtas POI grupes.
- [47] **Musisz** naładować baterię.
Reikia įkrauti akumuliatorių.
- [48] Język możesz zmienić później w Ustawieniach (**pagina** 81)
Vėliau ją galima keisti nuostatų ekrane (**puslapis** 79).
- [49] TMC (**Traffic Message Channel**)
TMC (**eismo pranešimų kanalas**)
- [50] **Wybierając** cel podróży trasy¹⁰.
Maršruto tikslą **pasirinkimas**.

W polskich zdaniach (przykłady [44]–[45]) zostaje zachowany właściwy dystans wobec użytkownika. Jednak w kolejnych przykładach (por. [46], [47], [48]) – typowych dla tej instrukcji – dystans ulega skróceniu, a w zdaniu [47] można wręcz mówić o niegrzeczności. W polskiej wersji tekstu (por. [48]) zostaje użyta forma sporadycznie stosowana na określenie strony. Dla przeciętnego odbiorcy może być nie zrozumiała. W zdaniu [49] nie przetłumaczono istotnego z punktu widzenia użytkownika sprzętu fragmentu instrukcji. W [50] błędnie zastosowano imiesłów przysłówkowy *wybierając* zamiast *wybór*. Tego typu błędy mogą świadczyć o niekompetencji tłumacza lub braku profesjonalizmu.

W litewskiej wersji instrukcji w całym tekście zostaje zachowany właściwy dystans wobec użytkownika – por. formy grzecznościowe w drugiej osobie liczby mnogiej (z charakterystyczną końcówką *-te*). W zdaniu [47] zastosowana neutralna stylistycznie konstrukcja *reikia įkrauti* odpowiada polskiej *należy naładować*. W przykładzie [48] przyjęty wariant litewski na określenie strony jest powszechnie stosowany, w [49] zaś – anglojęzyczny skrótowiec zostaje rozszyfrowany po litewsku. Również litewskojęzyczna nazwa podrozdziału (por. [50]) jest prawidłowa.

Zaprezentowane tu przykłady nie są wyjątkami. Niski poziom językowy polskich instrukcji obsługi sprzętów użytkowych to nie tylko efekt działań producenta czy importera sprzętu powszechnego użytku, to również przejaw ogólnego przyzwolenia na tego typu procedery, w szczególności – niewystarczający przejaw troski państwa o dobro, jakim jest język polski.

¹⁰ Jest to tytuł podrozdziału.

6. Podsumowanie

Eksperymentalny polsko-litewski korpus to rozbudowane narzędzie stworzone w Instytucie Sławistyki PAN przede wszystkim z myślą o badaniach konfrontatywnych i budowie planowanego polsko-litewskiego słownika elektronicznego. Potencjalnych odbiorców EKorpPL-LT (poza oczywistymi w tym względzie lingwistami) można upatrywać wśród informatyków (zajmujących się komputerowym przetwarzaniem języka), literaturoznawców, bibliotekoznawców, pedagogów, nauczycieli, tłumaczy, studentów, uczniów. Dwu- i wielojęzyczne korpusy są niezastąpione w procesie nauki języka obcego.

W artykule na przykładach kilku zagadnień ukazano możliwe sposoby zastosowania EKorpPL-LT. W toku przeprowadzanych eksperymentów okazało się, że korpus równoległy to nie tylko doskonałe narzędzie do konfrontacji, lecz również aparatura pozwalająca spojrzeć na każdy język z innej perspektywy. Bez tej właśnie perspektywy nie można byłoby tak wiele powiedzieć o polskich i litewskich zaimkach wskazujących, nieokreślonych, formach dualnych czy wołacza. Gdyby nie informacja uzyskana w oparciu o zasoby korpusowe, nikt nie byłby świadomy, że język polskich instrukcji sprzętów powszechnego użytku to nie tylko efekt komercji, lecz przede wszystkim wyraz słabości troski państwa o język polski.

Literatura

- Ambrazas V. (red.), 1997, *Dabartinės lietuvių kalbos gramatika* (Trečiasis pataisytas leidimas), Vilnius.
- Karolak S., 2001, *Od semantyki do gramatyki*, Warszawa.
- Koseska-Toševa V., Gargov G., 1990, *Българско-полска съпоставителна граматика*, t. I: *Фонетика и фонология*, София.
- Roszko D., 2013, *Eksperymentalny korpus litewskiej gwary puńskiej w Polsce (narzędzie do badań nad gwarami polskimi i litewskimi na Suwalszczyźnie)*, „Studia z Filologii Polskiej i Słowiańskiej” 48, s. 43–60.
- Roszko D., Roszko R., 2013, *Experimental Polish-Lithuanian corpus with the semantic annotation elements*, „Cognitive Studies | Études Cognitives” 13, s. 97–111; DOI: 10.11649/cs.2013.006.
- Roszko R., 2004, *Semantyczna kategoria określoności/nieokreśloności w języku litewskim (w zestawieniu z językiem polskim)*, Warszawa.
- Valeckienė A., 1998, *Funkcinė lietuvių kalbos gramatika*, Vilnius.

Applications of the experimental Polish-Lithuanian corpus /EkorpPL-LT/ in research on the contemporary Polish language and in theoretical Polish-Lithuanian contrastive studies

(Summary)

The experimental Polish-Lithuanian corpus is an extended tool put into effect at the Institute of Slavic Studies of the Polish Academy of Sciences, first of all with a view to contrastive studies and the structure of the planned Polish-Lithuanian electronic dictionary. The potential recipients of EKorpPL-LT (obviously apart from linguists) can comprise computer specialists (dealing with computer processing of a language), literary scholars, library science specialists, pedagogues, educationalists, teachers, translators, students. Bi- and multilingual corpora are irreplaceable and indispensable in the process of studying a foreign language.

In the article, basing on the examples of several problems, possible ways to apply EKorpPL-LT have been presented. During the experiments carried out it turned out that the parallel corpus is not only an excellent tool for comparison and contrast, but also an instrument allowing to look at every language from a different perspective. Without the perspective it would not be possible to say much about Polish and Lithuanian indicative pronouns, indefinite determiners, dual forms or the vocative case. If it were not for the information based on corpora resources, nobody would be aware that the language of Polish instructions of convenience appliances is not only a commercial effect, but above all an element of the state's concern for the Polish language.