Violetta Koseska-Toszeva
(emerytowana profesor Instytutu Sławistyki
Polskiej Akademii Nauk w Warszawie)

Roman Roszko
(Instytut Sławistyki Polskiej Akademii Nauk w Warszawie)

Języki słowiańskie i litewski w korpusach równoległych Clarin-PL

1. Wprowadzenie

1.1. Clarin-PL¹ włączył się do nurtu „Europejski jednolity rynek cyfrowy musi być wielojęzyczny!” Jak wiadomo, „strategia jednolitego rynku cyfrowego musi, dla zapewnienia równych szans cyfrowych we wszystkich językach urzędowych Unii Europejskiej, sprostać wyzwaniom wielojęzyczności”

---

¹ Międzynarodowa (europejska) infrastruktura naukowa Clarin ERIC (powstała w roku 2005, <http://www.clarin.eu/>). Jej częścią jest polskie konsorcjum Clarin-PL „Wspólne zasoby językowe i infrastruktura technologiczna” (<http://clarin-pl.eu/en/home-page/>), którego partnerami są Politechnika Wrocławska (lider konsorcjum, kierownik M. Piasecki), Instytut Podstaw Informatyki PAN, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, Instytut Sławistyki PAN, Uniwersytet Łódzki i Uniwersytet Wrocławski.

(Impact, 2009). Wciąż jednak główną przeszkodą na drodze do prawdziwie zjednoczonej europejskiej gospodarki i zjednoczonego europejskiego społeczeństwa pozostają bariery językowe. W celu pokonania tej przeszkody Clarin-PL opracowuje korpusy językowe języka polskiego w zestawieniu z innymi językami słowiańskimi oraz z językiem litewskim: *Korpus polsko-bułgarsko-rosyjski* (<http://hdl.handle.net/11321/308>) i *Korpus polsko-litewski* (<http://hdl.handle.net/11321/309>) (jest to zadanie Instytutu Slawistyki PAN). Korpusy te zapełniają lukę w słowiańskich i bałtosłowiańskich lingwistycznych zasobach elektronicznych i staną się ważnym krokiem ku tworzeniu miarodajnych przekładów tekstów zwykłych i elektronicznych. Jak wiadomo, powstały i nadal powstają (poza zadaniami Clarin-PL) korpusy wielojęzyczne, zawierające jako element składowy język polski. Są to między innymi korpusy polsko-ukraiński, polsko-słowacki, polsko-niemiecki, polsko-czeski, polsko-włoski, polsko-szwedzki, polsko-rosyjski UW, InterCorp, Opus, ParaSol, Emea, KDE4 i in. Szczególnie korpusy zawierające większą liczbę języków bazują na tekstach medycznych i unijnych (w mniejszym stopniu dotyczy to InterCorpu). Co należy podkreślić, w Clarin-PL powstał już korpus trzyjęzyczny, patrz niżej p. 2.1.

1.2. Strategicznym celem naukowym Clarin ERIC i Clarin-PL jest wspieranie badań humanistycznych w wielokulturowej i wielojęzycznej Europie. Wielokulturowość i wielojęzyczność pozostanie nadal nieodzownym elementem współczesnej Europy, potwierdzającym jej tożsamość. Dla Polski i polskich badaczy ważna jest budowa łącznika między językiem polskim i polskimi technologiami językowymi a innymi językami europejskimi i technologiami językowymi opracowanymi dla nich. Dotychczas w polskiej nauce największy nacisk był kładziony na powiązania polsko-angielskie. Tworzone przez Zespół Lingwistyki Korpusowej i Semantyki korpusy wnoszą nowe wartości i udostępniają nauce polskiej, polskiemu odbiorcy powiązane wielojęzyczne zasoby, w pierwszym etapie ograniczone do języków trzech grup Słowian oraz języka litewskiego ze względu na jego szczególne znaczenie dla polskiej historii i kultury oraz indoeuropejskiego językoznawstwa. Odbiorców obu powstających korpusów upatruje się wśród językoznawców, slawistów i bałtystów, w szczególności polonistów, rusycystów, bułgarystów i lituanistów, a także wśród leksykografów i terminologów, tłumaczy, osób nauczających i uczących się języków obcych, literaturoznawców, kulturoznawców, studentów różnych kierunków nauk humanistycznych, społecznych i technicznych (obszerna baza

tekstów specjalistycznych), przedstawicieli nauk stosowanych, inżynierów lingwistycznych, nauczycieli historii, wiedzy o świecie i sztuce.

1.3. Instytut Sławistyki PAN w Clarin-PL jest reprezentowany przez Zespół Lingwistyki Korpusowej i Semantyki (kierownik: Violetta Koseska-Toszewa, koordynator ze strony IS PAN do 30 IV 2016, Roman Roszko – przedstawiciel IS PAN w Radzie Konsorcjum od 1 V 2016), który w okresie 2013–2016 podjął się realizacji dwóch zadań, opisanych jako moduł 8 i moduł 9, por. niżej p. 2.

2. Korpusy IS PAN w Clarin-PL

W ramach polskiej części infrastruktury naukowej Clarin ERIC Zespół Lingwistyki Korpusowej i Semantyki opracowuje dwa korpusy równoległe: polsko-bułgarsko-rosyjski oraz polsko-litewski.

2.1. *Polsko-bułgarsko-rosyjski korpus równoległy* (moduł 8, odpowiedzialny: Wojciech Sosnowski, <http://hdl.handle.net/11321/308>) to wielojęzyczna baza wyselekcjonowanych, odpowiednio przygotowanych (format edytorski, standard zapisu) i zrównoleglonych tekstów w trzech wersjach językowych: polskiej, bułgarskiej i rosyjskiej. Objętość bazy jest określona na sześć milionów słowoform. Korpus zawiera współczesne teksty literackie, publicystyczne, naukowe i specjalistyczne. Wybór języków w korpusie jest nieprzypadkowy. Każda z grup języków słowiańskich posiada reprezentanta. Grupa języków zachodniosłowiańskich – język polski, grupa języków południowosłowiańskich – język bułgarski, grupa języków wschodniosłowiańskich – język rosyjski.

2.2. *Polsko-litewski korpus równoległy* (moduł 9, odpowiedzialny: Roman Roszko, <http://hdl.handle.net/11321/309>) to dwujęzyczna baza wyselekcjonowanych, odpowiednio przygotowanych i zrównoleglonych tekstów polskich i litewskich. Objętość bazy znacznie przekracza jedenaście milionów słowoform. Korpus zawiera współczesne teksty literackie, naukowe, publicystyczne, sądownicze, techniczne. W doborze języków do tego korpusu kierowano się spuścizną wielowiekowego wspólnego polsko-litewskiego dziedzictwa oraz archaicznym charakterem języka litewskiego.

2.3. Standard zapisu, objętość oraz dobór tekstów w obu korpusach spełnia podstawowe wymogi użyteczności baz wielojęzycznych do zastosowań

lingwistycznych, takich jak: prowadzenie ekscerpacji materiałowej, wydobywanie złożonych danych lingwistycznych, analiz konfrontatywnych, budowy słowników wielojęzycznych, ustalania strategii w procesie przekładu, wykrywania algorytmów międzyjęzykowych celem ich wykorzystania w przekładzie maszynowym i in.

2.4. Elementem **innowacyjnym** obu korpusów jest po raz pierwszy zdefiniowana i zastosowana anotacja semantyczna znaczeń kwantyfikacyjnych (por. niżej p. 3). Na pierwszym etapie prac nad korpusami założono tytułem eksperymentu badawczego ręczne naniesienie anotacji semantycznej na tekstach ciągłych do objętości dwóch tysięcy zdań dla każdego modułu.

3. Podstawy teoretyczne do anotacji semantycznej omawianych korpusów

3.1. Wprowadzanie do korpusów jedno- i wielojęzycznych elementów semantyki wykraczającej poza standardową morfosyntaktyczną anotację staje się standardem. Można tu wymienić chociażby webowy system do anotacji znaczeń IMI (Bond, Morgado da Costa, & Lê, 2015), w którym wielojęzyczne zasoby łączone są z bazami wordnetowymi celem ujednoznacznienia semantycznego leksemów. Elementy anotacji semantycznej odczytywane w nawiązaniu do FrameNetu zostały zaproponowane w pracy (Volk & Samuelsson, 2007) dla języków angielskiego i szwedzkiego, podobnie w pracy (Padó & Lapata, 2009), w której autorzy również opierają się na opisie ról semantycznych w korpusie angielsko-niemieckim w oparciu o (angielski) FrameNet. W Polsce natomiast w ramach zadań Clarin-PL powstaje zaawansowany słownik walencyjny języka polskiego („Walenty”, dostęp online: <http://walenty.ipipan.waw.pl/>), który zawiera informacje, w jaki sposób predykaty są powiązane z argumentami. „Walenty” jest przewidziany do głębokiego parsowania składniowego i składniowo-semantycznego. Można przytoczyć jeszcze kilka prac, które opisują zbliżone próby wzbogacenia anotacji korpusowej, korespondującej również z FrameNetem (por. Kingsbury, Palmer, & Marcus, 2002) oraz PropBankiem (por. Baker, Fillmore, & Lowe, 1998), np. projekt przedstawiony w Reeder i in. (2004). Proponowana tu anotacja semantyczna odbiega jednak od znanych w literaturze przedmiotu innych podejść do anotacji semantycznej, jako że abstrahując od semantyki leksykalnej i WordNetu czy FrameNetu opisuje

właściwe płaszczyźnie znaczeniowej treści odczytywane w oparciu o płaszczyznę formalną w powiązaniu z kontekstem, sytuacją i minimalną wspólną dla nadawcy i odbiorcy wiedzą o świecie (szczegóły: patrz niżej).

3.2. Anotacja semantyczna będąca przedmiotem analizy w tej pracy mieści się w semantyce z podejściem bezpośrednim (np. Russell, 1967). Wyróżniamy znaczenia kwantyfikacyjne imion i predykatów, w których wskazujemy znaczenia aspektowe i temporalne czasowników (np. kwantyfikacja pojedynczego przedmiotu: **(ix)P(x)** *ten znajomy*, kwantyfikacja pojedynczego zdarzenia: **(tX)P(X)** *(event_1) sprzedał*)². Opieramy się na logicznej kwantyfikacji zakresowej oraz na współczesnej teorii procesów, zwanej *sieciami Petriego*. Faktem godnym podkreślenia jest możliwość kwantyfikowania nie tylko imion, lecz także predykatu, co jest równoznaczne z kwantyfikacją czasu i aspektu. Co bardzo ważne, nie utożsamiamy kategorii aspektu z rodzajem czynności, co może sprawdzać się ewentualnie w tych językach, w których nie ma rozwiniętej kategorii aspektu (tak jak w językach słowiańskich). Semantyka sytuacyjna, do czego przyznają się w swych pracach Jon Barwise i John Perry, jest bliska zarówno ideom Bertranda Russella i Andrzeja Stanisława Mostowskiego, jak i intuicji lingwistów, zwłaszcza tych, którzy zajmowali się gramatyką funkcjonalną (por. Barwise & Perry, 1983). Teoria sieci Petriego – ważna dla określenia czasu i aspektu – jest także teorią z bezpośrednim podejściem do semantyki języka naturalnego (zob. Petri, 1962; Mazurkiewicz, 1986; Koseska-Toszewa & Mazurkiewicz, 1988, 1994; Mazurkiewicz & Koseska-Toszewa, 1991; R. Roszko, 2004).

3.3. Sieci Petriego zbudowane są ze skończonej liczby obiektów symbolizujących stany i zdarzenia, połączonych relacją następstwa. Relacje następstwa prowadzą albo od zdarzenia do stanu (gdy zdarzenie rozpoczyna stan), albo od stanu do zdarzenia (gdy stan jest zakończony zdarzeniem). Relacja następstwa nie musi być porządkiem liniowym; pewne obiekty sieci mogą być nieporównywalne pod względem porządku, gdy żadnego z nich nie poprzedza inne. Przez **stan** rozumiemy własność pewnego obiektu rzeczywistości. W dyskretnym podejściu do opisu procesów paradygmatem stanu jest jego **trwanie**. Każdy stan trwa jakiś określony czas. Dwa różne stany następujące po sobie dzieli jakieś zdarzenie, rozpoczynające nowy stan i kończące stary.

---

² W stosowanym przez nas zapisie „x” odnosi się do przedmiotów, natomiast „X” – do stanów i zdarzeń.

Uogólniając, **zdarzenie** kończy istnienie pewnego stanu i/lub rozpoczyna istnienie innego. Może być traktowane jako punkt na osi czasu, gdyż będąc granicą między stanami, nie może być rozciągle w czasie (nie trwa). Na przykład cztery pory roku są stanami; równonoc i przesilenia słoneczne są zdarzeniami, np. wiosenna równonoc (zdarzenie) oddziela zimę (stan) od wiosny (stanu), zob. Petri (1962); Mazurkiewicz (1986).

Rozróżnienie stanów i zdarzeń jest istotną cechą sieci Petriego. Każde zdarzenie kończy albo zaczyna stan; dwa różne stany następujące po sobie muszą być rozdzielone przez jakieś zdarzenie kończące jeden z nich i zaczynające następny. Podobnie między dwoma następującymi po sobie zdarzeniami zawsze istnieje stan (dający się opisać chociażby tak: „pierwsze zdarzenie już zaszło, a drugie jeszcze nie nastąpiło”). Typowymi znaczeniami form dokonanych czasownika są: zdarzenie (event_1) i szereg stanów i zdarzeń zakończonych zdarzeniem (event_2). Z kolei niedokonanej formie verbum przypisane są znaczenia: stanu (state_1) i szeregu stanów i zdarzeń zakończonych stanem (state_2). Znaczenia zdarzeń i stanów można dobrze zilustrować na przykładzie relacji aspektowo-temporalnej, czyli wówczas, kiedy forma werbalna wyraża określony czas w zdaniu.

Semantyczna kategoria czasu to kategoria porządkująca stany i zdarzenia względem stanu wypowiedzi za pomocą relacji uprzedniości-następczości (Koseska-Toszewa, 2006). Z przyjętego postulatu skończoności modelu wynika, że w opisie nie możemy ograniczyć się jedynie do zdarzeń, a w konsekwencji traktować stanów jako zbiorów zdarzeń, tak jak to robił między innymi Reichenbach (1967). Przy opisie stanu jako zbioru zdarzeń nasuwają się bowiem pytania: których zdarzeń? Wszystkich czy niektórych? A jeśli niektórych, to jak je wybrać? Pominięcie zdarzeń w modelu i ograniczenie się wyłącznie do stanów pozbawia nas możliwości rozpatrywania zjawisk, takich jak „zderzenie”, „otwarcie”, „odsłonięcie”, „przebudzenie” i tym podobnych.

3.4. Semantyczna kategoria określoności/nieokreśloności. Badania nad tradycyjnie rozumianą kategorią określoności/nieokreśloności sprowadzały się zazwyczaj do opisu jej wykładników morfologicznych przede wszystkim w tzw. językach rodzajnikowych.

„Natomiast w językach bezrodzajnikowych – jak pisze Laskowski (2011, ss. 172–176) – szukano leksykalnych odpowiedników odpowiadających treściom przypisywanym użyciu rodzajnika. W związku z powyższym kategorię określoności/nieokreśloności traktowano wyłącznie jako kategorię frazy nominalnej.

Powyższy fakt długie lata ciążył na opisach tej kategorii, często sprowadzanych wyłącznie do analizy znaczeń zaimków w językach bezrodzajnikowych. [...] Wspomniany tom 2. autorstwa V. Koseskiej-Toszewej i G. Gargova (wydany w Sofii), odwołujący się do teorii deskrypcji Bertranda Russella i semantyki sytuacyjnej J. Barwise'a i J. Perry'ego oraz R. Coopera (uwzględniającej sytuację komunikacji językowej oraz stany jej uczestników i ich stosunek do komunikowanej treści) jako semantycznego punktu wyjścia do analizy materiału językowego, jest dziełem pionierskim nie tylko ze względu na przyjęty tam rachunek kwantyfikatorów jako metajęzyk opisu kategorii określoności. Tom przynosi nową interpretację tej kategorii w obu językach, obejmującą – co stanowiło wówczas istotne novum – zarówno kwantyfikację grup imiennych, jak i fraz werbalnych (kwantyfikacja stanów i zdarzeń, czasu)”.

W Polsce nie był dostatecznie znany kwantyfikacyjny model opisu kategorii określoności/nieokreśloności, choć na świecie właśnie zakresowa kwantyfikacja logiczna była już wówczas uznana za odpowiedni model teoretyczny, uściślający pojęcia związane z tą kategorią językową (por. Barwise & Cooper, 1981, ss. 159–219; Bellert, 1971; Cooper, 1996; Desclés, 1999; R. Roszko, 2004).

3.4.1. Badania Koseskiej-Toszewej (1982), Koseskiej-Toszewej, Gargova (Kocecka-Tomeba & Габров, 1990), R. Roszko (2004) i D. Roszko (2015) pokazały, że kategoria określoności/nieokreśloności jako kategoria semantyczna (a) jest wyrażana różnymi środkami leksykalnymi i morfologicznymi na poziomie fraz nominalnej i werbalnej, (b) jest kategorią zdania (a nie tylko kategorią frazy nominalnej!). Używanie terminu „określoność” w przypadkach, kiedy tzw. „rodzajnik określony” wyraża nieokreśloność, czyli ogólność, było oczywistym błędem i wynikało z nierozróżnienia formy od jej znaczenia. W pracach Koseskiej-Toszewej (1982) oraz Koseskiej-Toszewej i Gargova (Kocecka-Tomeba & Габров, 1990) kategoria określoności/nieokreśloności została zdefiniowana jako kategoria z semantyczną opozycją jednostkowości do niejednostkowości, przy czym przez określoność rozumie się wyłącznie jednostkowość elementu lub zbioru (spełniającego predykat), a przez nieokreśloność – niejednostkowość (obejmującą znaczenia egzystencjalności i ogólności), por. również (Koseska-Toszewa, Korytkowska, & R. Roszko, 2007; R. Roszko, 2004; D. Roszko, 2015).

W języku bułgarskim za najbardziej typowy środek morfologiczny wyrażający we frazie nominalnej jednostkowość (określoność) i ogólność (nieokreśloność) uważa się rodzajnik (*-ѣm/-a*, *-ma*, *-mo*). Jego brak, czyli morfologiczne zero ($\emptyset$), jest znaczący – jest on bowiem wykładnikiem egzystencjalności albo czystej predykcji. Dwuznaczność rodzajnika bułgarskiego stanowi dobrą

ilustrację trudności, z jakimi styka się badacz tej kategorii przy klasyfikacji kwantyfikacyjnej wyrażen z języka naturalnego. Analogiczne zjawisko jest obserwowane także w języku litewskim (por. lit. *mylimasis* ‘ten ukochany’ (jednostkowość) i *mylimasis* ‘każdy kto kocha’ (ogólność)).

3.4.2. Przypomnijmy, że wykorzystując ogólnie przyjęte definicje kwantyfikatorów logicznych i operatora jota posługujemy się trzema podstawowymi pojęciami, na których znaczenie składają się językowe wykładniki kwantyfikacji logicznej: jednostkowej, egzystencjalnej i ogólnej (Rasiowa, 1975, ss. 211–255) i deskrypcji określonej (Russell, 1967, ss. 253–293), por. Koseska-Toszewa (1982); Kocicka-Toшева & Габров (1990).

3.4.3. W literaturze przedmiotu morfologiczne środki wyrażania kategorii określoności we frazie werbalnej w obu językach nie były dostatecznie badane, por. wyjątkowo (Grzegorczykowa, 1976; Koseska-Toszewa, 1982; R. Roszko, 2004; D. Roszko & R. Roszko, 2007) i ostatnio (D. Roszko, 2015). Na poziomie frazy werbalnej kwantyfikuje się przede wszystkim czas i aspekt oraz miejsce trwającego stanu, czy zajścia zdarzenia. Wielu autorom kwantyfikacja zakresowa kojarzyłaby się tylko z aspektem. Jednak aoryst bułgarski, niezależnie od informacji o aspekcie, rezerwuje miejsce dla kwantyfikatora jednostkowości. Jednostkowość kwantyfikacyjna wyrażana jest formami aorystu utworzonego zarówno od czasowników dokonanych, jak i niedokonanych. W bułgarskim żadna z form aorystu nie występuje obok wyrażenia kwantyfikującego egzystencjalnego czy też uniwersalnego (Koseska-Toszewa, 2006). Dzisiaj model oparty na logicznej teorii kwantyfikacji w opisie kategorii określoności/nieokreśloności jest coraz częściej wykorzystywany w świecie (por. Barwise & Cooper, 1981; Bellert, 1971; Grzegorczykowa, 1972, 1976; Kocicka-Toшева & Габров, 1990; Cooper, 1996; Desclés, 1999; R. Roszko, 2004; D. Roszko, 2015). Najważniejsze jest odnalezienie środków językowych wyrażających kwantyfikację. Nie jest to zadania łatwe, np. dla Grzegorczykowej wyraz *niektórzy* oznacza ogólność o ograniczonym zakresie, Koseska-Toszewa zaś rozpatruje to wyrażenie jako wyrażenie kwantyfikacji egzystencjalnej.

3.4.4. W wielu pracach znaczenie aspektu rozpatrywane jest w ścisłej łączności ze znaczeniem temporalnym verbum (por. Иванчев, 1971). Do łącznego rozpatrywania znaczenia aspektu i czasu upoważnia zwłaszcza materiał z języka bułgarskiego. W pracach nad aspektem trwają dyskusje wokół sposobów kształtowania się jego postaci formalnej. Najczęściej mówi się o znaczeniach aspektu jako o dokonaności i niedokonaności czynności. Jednak co to znaczy?

Dyskusje na ten temat toczą się m.in. w teoriach Andrzeja Bogusławskiego (2003) i Stanisława Karolaka (2008). Teoria matematyczna sieci Petriego daje odpowiedź na pytanie o znaczenia aspektu i czasu w językach słowiańskich, a także o znaczenia czasu w innych językach naturalnych, w których nie wykształciła się formalna kategoria aspektu. Pojęcia zdarzeń, stanów i ich konfiguracji rozumiane są tu tak jak w opisie sieciowym czasu i aspektu (por. szczegółowiej nt. zastosowania teorii sieci Petriego w języku naturalnym: Mazurkiewicz, 1986; Koseska-Toszeza & Mazurkiewicz, 1988, 2010; Koseska-Toszeza, 1995, 1997, 2006; R. Roszko, 2004; D. Roszko & R. Roszko, 2014).

3.4.5. Niedopowiedzenie kwantyfikacji. Jeśli przyjmiemy za Frege (1982), że prawda i prawdziwość zdania to spełnianie pewnego warunku w zdaniu w sensie logicznym, powinniśmy sobie zdać sprawę z tego, że mówimy nie o zdaniu w sensie syntaktycznym i formalnym, lecz o jego semantycznej strukturze³. W terminologii Ajdukiewicza (1965) mamy do czynienia z sądem i zdaniem: „[...] jakieś wyrażenie nazwiemy przy pewnym jego znaczeniu zdaniem oznajmującym (lub zdaniem w sensie logicznym), jeżeli wyrażenie to wypowiada przy tym właśnie znaczeniu jakiś sąd, tzn. jakąś myśl odnoszącą się w sposób sprawozdawczy do jakiegoś stanu rzeczy”. Zdania oznajmujące Ajdukiewicz nazywa zdaniami w sensie logicznym. Według niego mają one pewną wspólną własność odróżniającą je od wszystkich pozostałych rodzajów zdań. Tą cechą wszystkich zdań oznajmujących (i tylko tych!) jest mianowicie to, że są albo prawdą, albo fałszem. W języku naturalnym spotykamy często zdania oznajmujące bez wartości prawdy i fałszu, takie jak przytaczane przez Ajdukiewicza (1965) zdanie: *Anglicy są flegmatyczni*. Ajdukiewicz zwraca uwagę na fakt, że to zdanie jest przykładem niedopowiedzenia kwantyfikacji, i dlatego nie ma wartości ani prawdy, ani – odpowiednio – fałszu. Nie wiemy bowiem, czy chodzi tu o wszystkich Anglików, o ich większość, o niektórych, czy tylko o dokładnie tych Anglików. Mówiąc w sposób uproszczony, nadawca tej informacji nie wie, czy predykat *X są flegmatyczni* jest spełniony dla nazwy „Anglicy”. Potrzebna jest w danym wypadku dodatkowa informacja mogąca nadać temu predykatowi wartość prawdy. Mianowicie brakuje tu informacji o kwantyfikacji zakresowej nazwy „Anglicy”. To zjawisko zostało zauważone przez Ajdukiewicza i opisane jako niedopowiedzenie kwantyfikacji wraz

---

³ Świadomie unikamy tu pojęć sądu i propozycji, ponieważ są to pojęcia różnorako definiowane w literaturze przedmiotu.

z niedopowiedzeniem relatywizacji w zdaniach z języka naturalnego. W języku polskim częściej niż w bułgarskim czy litewskim stykamy się z takim zjawiskiem. Ajdukiewicz (1965) mówi o niedopowiedzeniu kwantyfikacji, mając na uwadze tylko frazę nominalną, natomiast my będziemy tu mówić o niedopowiedzeniu logicznej kwantyfikacji zakresowej, mając na uwadze także frazę werbalną w zdaniach zestawianych języków.

3.4.5.1. Niedopowiedzenie kwantyfikacji możemy zarejestrować jednocześnie w czterech tu omawianych językach, por. przykład niedopowiedzenia kwantyfikacji egzystencjalnej na poziomie frazy nominalnej:

PL	<i>Na dworze płacze <u>dziecko</u>.</i>	Niedopowiedziana egzystencjalność – <i>jakiś dziecko</i>
BG	<i>На двора плаче <u>дете</u>.</i>	Niedopowiedziana egzystencjalność – <i>едно, някакво дете</i>
RU	<i>Во дворе плачет <u>ребенок</u>.</i>	Niedopowiedziana egzystencjalność – <i>какой-то ребенок</i>
LT	<i>Lauke verkia <u>vaikas</u>.</i>	Niedopowiedziana egzystencjalność – <i>kažkoks vaikas</i>

Możliwe są również przykłady braku jednoczesnego niedopowiedzenia kwantyfikacji w omawianych językach, por. niżej zdania polskie, rosyjskie i litewskie z niedopowiedzianą kwantyfikacją ogólną stojące w kontraście do zdania bułgarskiego:

PL	<i>Praca ma największą wartość.</i>	Niedopowiedziana ogólność – <i>każda praca</i>
BG	<i>Трудът е с най-висока стойност.</i>	Ogólność jest wyrażona jednym ze znaczeń rodzajnika - <i>ът</i>
LT	<i>Darbas yra didžiausia vertė.</i>	Niedopowiedziana ogólność – <i>kiekvienas darbas</i>
RU	<i>Труд – это наибольшая ценность.</i>	Niedopowiedziana ogólność – <i>каждый труд</i>

Na poziomie frazy werbalnej niedopowiedzenie może dotyczyć egzystencjalnej, ogólnej i jednostkowej kwantyfikacji. W rozpatrywanych językach zazwyczaj są to zdania bez wyrażeń kwantyfikujących typu: bułg. *винаги, от време на време, понякога*, pol. *zawsze, niekiedy, czasami*, ros. *всегда, никогда*, czy lit. *visada, kartais*, por. p. 4.3. Należy jednak podkreślić, że niedopowiedzenie kwantyfikacji we frazie werbalnej jest najrzadziej rejestrowane w tekście litewskim za sprawą między innymi jednoznacznych kwantyfikacyjnie form tzw. czasu przeszłego wielokrotnego, form czasu teraźniejszego utworzonych od czasowników momentalnych oraz zdecydowanie wyższego nasycenia zdań formami powiązаныmi z czasownikiem i imiesłowem jednoznacznie wskazującymi na rodzaj kwantyfikacji frazy werbalnej.

4. Typy anotacji semantycznej zastosowanej po raz pierwszy w omawianych korpusach

W tej części zostają zaprezentowane poszczególne znaczenia kwantyfikacyjne w rozbiciu na frazę nominalną (p. 4.1) i werbalną (p. 4.2)⁴.

4.1. Anotacja frazy nominalnej. Przykłady poszczególnych znaczeń kwantyfikacyjnych we frazie nominalnej⁵:

$\langle \text{ix} \rangle \text{P}(\text{x})$ – kwantyfikator jednostkowości dotyczący frazy nominalnej (jednostkowość przedmiotu)

PL	- Wróć z powrotem – cicho odparł $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Rimski</i> \> – ale $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Lichodiejew</i> \> ciężko odpokutuje $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>ten piknik</i> \>.
BG	- Ще $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>ѐи</i> \> върнат – обади се тихо $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Римски</i> \>, – обаче $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>на него</i> \> $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>това излетче</i> \> ще $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>му</i> \> излезе през носа.
RU	- $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Они</i> \> придут обратно, – отозвался $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Римский</i> \> тихо, – а вот $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>он</i> \> сильно ответит $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>за этот пикничок</i> \>.
LT	<i>Pinigai grįš</i> , – tyliu balsu atsakė $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>Rimskis</i> \>, – o štai $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>jam</i> \> $\langle \text{ix} \rangle \text{P}(\text{x})$ <i>tasai piknikėlis</i> \> smarkiai atsirūgs.

$\langle \exists \text{x} \rangle \text{P}(\text{x})$ – kwantyfikator egzystencjalności dotyczący frazy nominalnej (egzystencjalność przedmiotu)

PL	<i>Na balkonie</i> $\langle \exists \text{x} \rangle \text{P}(\text{x})$ <i>jakiś wstrząśnięty obywatel</i> \> znalazł paczkę banknotów w bankowej banderoli z napisem: „Jeden tysiąc rubli”.
BG	<i>На втори балкон</i> $\langle \exists \text{x} \rangle \text{P}(\text{x})$ <i>някакъв слисан гражданин</i> \> откри в джоба си пачка с банков бандерол с надпис „Хиляда рубли”.
RU	<i>На галерее</i> $\langle \exists \text{x} \rangle \text{P}(\text{x})$ <i>какой-то смятенный гражданин</i> \> обнаружил у себя в кармане пачку, перевязанную банковским способом и с надписью на обложке: «Одна тысяча рублей».
LT	<i>Balkone</i> $\langle \exists \text{x} \rangle \text{P}(\text{x})$ <i>kažkoks sutrikęs pilietis</i> \> rado savo kišenėje perjuostą, lyg tiesiai iš banko, pluoštelį su užrašu: „Vienas tūkstantis rublių”.

⁴ Prezentowane tu cytaty pochodzą z korpusów równoległych polsko-bułgarsko-rosyjskiego i polsko-litewskiego, powstających w ramach zadań Clarin-PL.

⁵ Znak „<” oznacza początek opisu, znak „>” oznacza koniec opisu. Po symbolu „<” podawany jest rodzaj kwantyfikacji, następnie spacja i definiowane wyrażenie imienne.

$(\forall x)P(x)$ – kwantyfikikator ogólności dotyczący frazy nominalnej (ogólność przedmiotu)

PL	$\langle(\forall x)P(x)$ <i>Każdy, kto trafił do tego domu \></i> , jeśli, oczywista, nie był zupełnym tępakiem, od razu widział, jak dobrze się żyje wybrancom losu – członkom Massolitu, i z miejsca opanowywała go czarna zawiść.
BG	<i>Щом се озовеше в Грибоедов, $\langle(\forall x)P(x)$ всеки посетител \></i> , стига да не беше абсолютен тъпак, разбира се, веднага схващаше колко хубаво си живеят щастливците – членове на МАССОЛИТ, и черна завист почваше незабавно да го гложди.
RU	$\langle(\forall x)P(x)$ <i>Всякий посетитель \></i> , если он, конечно, был не вовсе тупицей, попал в Грибоедова, сразу же соображал, насколько хорошо живет ся счастливым – членам МАССОЛИТа, и черная зависть начинала немедленно терзать его.
LT	$\langle(\forall x)P(x)$ <i>Kiekvienas lankytojas \></i> , jeigu tik jis, žinoma, nebuvo visai bukagalvis, pakliuvęs į Gribojedovą, beregint suvokdavo, kaip šauniai gyvena tie laimės kūdikiai – MASSOLIT'o nariai, ir tučtuojau ji imdavo graužti siaubingas pavydas.

4.2. Anotacja frazy werbalnej. Kwantyfikacja dotycząca czasu i aspektu odnosi się do (a) zdarzeń (event_1) lub (b) stanów (state_1) oraz (c) szeregów stanów i zdarzeń zakończonych zdarzeniem (event_2) i (d) szeregu stanów i zdarzeń zakończonych stanem (state_2). Por. niżej przykład, w którym kwantyfikacja egzystencjalna odnosi się do szeregu stanów i zdarzeń zakończonych stanem:

PL	$\langle(\exists x)P(x)$ <i>(state_2) Był sobie \></i> krasnoludek z długą białą brodą, w czerwonej czapce.
BG	$\langle(\exists x)P(x)$ <i>(state_2) Имаше някога \></i> едно джудже с дълга бяла брадица и алена шапчица.
RU	$\langle(\exists x)P(x)$ <i>(state_2) Жил некогда \></i> на свете гном с длинной белой бородой.
LT	$\langle(\exists x)P(x)$ <i>(state_2) Gyveno \></i> raudonkepuraitis nykštukas su ilga balta barzda.

Przykłady poszczególnych znaczeń kwantyfikacyjnych we frazie werbalnej:

$(\iota x)P(X)$ – kwantyfikikator jednostkowości dotyczący frazy werbalnej (jednostkowość zdarzenia)

PL	<i>Piłat... – $\langle(\iota x)P(X)$ w tym momencie zamknął \></i> oczy.
BG	<i>Пилат... – и $\langle(\iota x)P(X)$ затвори \></i> очи.
RU	<i>Пилат... – тут он $\langle(\iota x)P(X)$ закрыл \></i> глаза.
LT	<i>Pilotas... – ir $\langle(\iota x)P(X)$ užsimerkė \></i> .

$(\exists X)P(X)$ – kwantyfikator egzystencjalności dotyczący frazy werbalnej (podawany przykład egzystencjalności szeregu stanów i zdarzeń zakończonych stanem)

PL	$\langle (\exists X)P(X)_{\text{state\_2}} \text{ Czasami chował się } \rangle$ do bram, $\langle (\exists X)P(X)_{\text{state\_2}} \text{ unikał } \rangle$ skrzyżowań ze światłami i eleganckich podjazdów przed willami dyplomatów.
BG	$\langle (\exists X)P(X)_{\text{state\_2}} \text{ От време на време се криеше } \rangle$ във входовете и $\langle (\exists X)P(X)_{\text{state\_2}} \text{ избягваше } \rangle$ кръстопътищата със светофари и разкошните врати на посолствата.
RU	$\langle (\exists X)P(X)_{\text{state\_2}} \text{ По временам прячется } \rangle$ в подъездах и $\langle (\exists X)P(X)_{\text{state\_2}} \text{ избегая } \rangle$ перекрестков со светофорами, шикарных дверей посольских особняков.
LT	$\langle (\exists X)P(X)_{\text{state\_2}} \text{ Tarpais sprukdamas } \rangle$ į tarpuvartes ir $\langle (\exists X)P(X)_{\text{state\_2}} \text{ aplenkdamas } \rangle$ sankryžas su šviesoforais ir prašmatnias pasiuntinybių duris.

$(\forall X)P(X)$ – kwantyfikator ogólności dotyczący frazy werbalnej (podawany przykład ogólności szeregu stanów i zdarzeń zakończonych stanem)

PL	$\langle (\forall X)P(X)_{\text{state\_2}} \text{ Nigdy niczego podobnego nawet w ręku nie miałem } \rangle$! – rozpaczliwie krzyczał prezes.
BG	Такова нещо в ръце $\langle (\forall X)P(X)_{\text{state\_2}} \text{ не съм хващал } \rangle$! – извика отчаяно председателят.
RU	$\langle (\forall X)P(X)_{\text{state\_2}} \text{ Никогда в руках не держал } \rangle$! – отчаянно вскричал председатель.
LT	$\langle (\forall X)P(X)_{\text{state\_2}} \text{ Niekada nė į rankas nebuvaui paėmęs } \rangle$! – beviltiškai suklykė pirmininkas.

4.3. Niedopowiedzenie kwantyfikacji

4.3.1. Przykłady poszczególnych niedopowiedzianych znaczeń kwantyfikacyjnych we frazie nominalnej⁶:

$?(\text{ix})P(\text{x})$ – niedopowiedzenie jednostkowości przedmiotu

PL	$\langle ?(\text{ix})P(\text{x}) \text{ Dziewczyna } \rangle$ była bardzo młoda.
BG	Девојката беше много млада. (Uwaga! W języku bułgarskim w wyniku znaczenia formy rodzajnika nie ma niedopowiedzenia kwantyfikacji. Forma rodzajnika oznacza tu jednostkowość przedmiotu)
RU	$\langle ?(\text{ix})P(\text{x}) \text{ Девушка } \rangle$ была очень молодой.
LT	$\langle ?(\text{ix})P(\text{x}) \text{ Mergaitė } \rangle$ buvo labai jauna.

⁶ Znak zapytania „?” w prepozycji formuły informuje o niedopowiedzianej kwantyfikacji.

?($\exists x$)P(x) – niedopowiedzenie egzystencjalności przedmiotu

PL	Zadzwo <i>ni</i> ła <?($\exists x$)P(x) dziewczyna \>. (jakaś)
BG	Обади се <?($\exists x$)P(x) девојка \>. (някаква)
RU	<?($\exists x$)P(x) Девушка \> позвонила. (какая-то)
LT	<?($\exists x$)P(x) Mergaitė \> paskambino. (kažkokia)

?($\forall x$)P(x) – niedopowiedzenie ogólności przedmiotu

PL	<?($\forall x$)P(x) Człowiek \> jest śmiertelny.
BG	Човекът е смъртен. (Uwaga! W języku bułgarskim w wyniku znaczenia formy rodzajnika nie ma niedopowiedzenia kwantyfikacji)
RU	<?($\forall x$)P(x) Человек \> смертельный.
LT	<?($\forall x$)P(x) Žmogus \> yra mirtingas.

4.3.2. Przykłady poszczególnych niedopowiedzianych znaczeń kwantyfikacyjnych we frazie werbalnej:

?(ιx)P(X) – niedopowiedzenie jednostkowości frazy werbalnej

PL	Maciej <?(ιx)P(X) <u>otwiera</u> \> okno. (Nie wiadomo, czy teraz, czy czasami, czy zawsze)
BG	Матеј <?(ιx)P(X) <u>отваря</u> \> прозореца. (jw.)
RU	Он <?(ιx)P(X) <u>открывает</u> \> окно. (jw.)
LT	Jis <?(ιx)P(X) <u>atidaro</u> \> langą. (jw.)

?($\exists X$)P(X) – niedopowiedzenie egzystencjalności frazy werbalnej

PL	Dziewczyna <?($\exists X$)P(X) spoglądała \> w jego stronę. (Nie wiadomo, czy wtedy, czy od czasu do czasu, czy zawsze)
BG	Девојката <?($\exists X$)P(X) поглеждаше \> в негова посока. (jw.)
RU	Девушка <?($\exists X$)P(X) смотрела \> в его сторону. (jw.)
LT	Mergaitė <?($\exists X$)P(X) žvelgė \> jo link. (jw.)

?($\forall X$)P(X) – niedopowiedzenie ogólności frazy werbalnej/predykatu

PL	On <?($\forall X$)P(X) się spóźnia \>. (Nie wiadomo, czy zawsze, czy czasami)
BG	Той <?($\forall X$)P(X) закъснява \>. (jw.)
RU	Он <?($\forall X$)P(X) опаздывает \>. (jw.)
LT	Jis <?($\forall X$)P(X) vėluoja \>. (jw.)

4.4. Anotacja semantyczna będzie zastosowana w tekstach ciągłych z korpusów równoległych Clarin-PL. Tekst ciągły zmienia bowiem kwantyfikację imion, por. na przykładzie z języka polskiego (forma *krasnodudek*):

<(∃x)P(x)_(state_2) Był sobie \> <?(∃x)(P(x) *krasnodudek* \> z długą białą brodą, w czerwonej czapce. [...] Kilka mądrych słów <(ix)P(x) *krasnodudka* \> <(ix)P(x)_(state_2) wystarczyło \>, by zażegnać kłótnie między mrówką a konikiem polnym.

Jak widać w dalszej części tekstu kwantyfikacja dotycząca *krasnodudka* staje się jednostkowa.

4.5. Nanoszenie anotacji jest zadaniem trudnym, wymagającym od anotatora nie tylko doskonałej znajomości danych języków, lecz także doskonałego zrozumienia teorii, które legły u podstaw wyróżnienia poszczególnych znaczeń kwantyfikacyjnych, oraz doświadczenia w analizie płaszczyzny semantycznej zdania. Aby choć w części ukazać te trudności, podajmy przykłady z czterech języków.

Lp.	Język polski	Język bułgarski	Język rosyjski	Język litewski
1.	Powodem tego była z pewnością jak zwykle krew, która napłynęła do skroni i łomotała w nich teraz, ale <i>coś</i> jednocześnie stało się ze wzrokiem prokuratora.	Пак трябва да беше от кръвта, нахлула към слепоочията и затуптяла в тях, но <i>нещо</i> се случи и със зрението му.	Опять-таки вино-вата была, вероятно, кровь, прилившая к вискам и застучавшая в них, только у прокуратора <i>что-то</i> случилось со зрением.	Matyt, dėl to paties kraujo, suplūdusio į smilkinius ir ėmusio tvinkčioti juose, <i>kažkas</i> atsitiko ir prokuratoriaus regėjimui.
2.	Kilkakrotnie próbował wstać, ale nogi odmawiały mu posłuszeństwa – Bezdomnemu przydarzyło się <i>coś</i> , co można porównać jedynie do paraliżu.	Няколко пъти понечи да се изправи, но краката не го слушаха – беше получил <i>нещо</i> като парализа.	Несколько раз он пытался подняться, но ноги его не слушались – с Бездомным приключилось <i>что-то</i> вроде паралича.	Keletą kartų jis bandė keltis, bet kojos neklausė – Benamį ištiko <i>kažkas</i> panašaus į paralyžį.

Lp.	Język polski	Język bułgarski	Język rosyjski	Język litewski
3.	Przejęci ludzie biegli alejką obok poety, wołali coś , ale do Iwana ich słowa nie docierały.	Развълнувани хора тичаха край поета по алеята, възклицаваха нещо , но Иван Николаевич не възприемаше думите им.	Взволнованные люди пробегали мимо поэта по аллее, что-то восклицая, но Иван Николаевич их слов не воспринимал.	Susijaudinę žmonės vis bėgo aleja pro šalį kažką šūkčiodami, tačiau Ivanas Nikolajevičius nesuprato jų žodžių.
4.	Czy kiedykolwiek mówiłeś coś o wielkim Cezarze?	Говорил ли си някога нещо за великия кесар?	Ты когда-либо говорил что-нибудь о великом кесаре?	Ar tu kada nors kalbėjai ką nors apie didįjį cezarių?
5.	– Ale potrzebne są jakiś na to świadectwa ... – zaczął Berlioz.	– Но нали трябва да има някакво доказателство... – отвори уста Берлиоз.	– Но требуется же какое-нибудь доказательство... – начал Берлиоз.	Bet juk reikia kokių nors įrodymų ... – buvo bepradedęs Berliozas.
6.	Kiedyś przy kieliszku opowiem panu kilka faktów z mojego życia, dając słowo – nie powstrzyma się pan od śmiechu!	Някой ден на чашка ще ви разкажа такива факти от моята биография, че ще си умрете от смях!	Как-нибудь за рюмкой я вам расскажу несколько фактов из моей биографии, вы обхотечесь!	Kada nors prie taurelės papasakosiu jums keletą nutikimų iš savo biografijos, leipsite iš juoko!

Z analizy płaszczyzny semantycznej przykładów 1–3 wynika, że wszystkie wyróżnione formy są wykładnikami znaczeń kwantyfikacyjnych egzystencjalnych. Przy okazji można wskazać prostą formalną zależność między wyróżnionymi leksemami z trzech języków, mianowicie polskiemu *coś* (zawierającemu formant *-ś*) odpowiada rosyjskie *что-то* (zawierające formant *-то*) i litewskie *kažkas* (zawierające formant *kaž-*). Uogólniając ukazane w przykładach 1–3 zjawisko stwierdza się, że polskiej serii zaimków z formantem *-ś* regularnie odpowiadają: rosyjska seria zaimków z formantem *-то* i litewska z formantem *kaž-*; por. przykłady z innym zaimkiem:

pol. *I wtedy w mózgu Berlioza **ktoś** rozpaczliwie krzyknął...*

bułg. *В мозъка на Берлиоз **някой** отчаяно изкрещя...*

ros. *Тут в мозгу Берлиоза **кто-то** отчаянно крикнул...*

lit. *Tą akimirksnį Berliozo smegenyse **kažkas** beviltiškai suriko.*

W przykładach 4–6 znaczenie kwantyfikacyjne wyróżnionych form określa się jako ogólność. Jak można zauważyć, w języku polskim nadal pojawiają się formy z formantem -ś, jednak – zwróćmy na to uwagę – w litewskim i rosyjskim rejestrujemy zupełnie inną serię zaimków: w rosyjskim z formantem *-нибудь* (*-либо*), w litewskim z formantem *nors*. Oznacza to, że języki rosyjski i litewski jednoznacznie (i zawsze) wyróżniają znaczenia ogólności na płaszczyźnie formalnej, czego nie można powiedzieć o językach bułgarskim i polskim. Zauważmy: tylko w jednym zdaniu (4) polski formant -ś zostaje zastąpiony jednoznacznym formantem kwantyfikacyjnych znaczeń ogólności *-kolwiek*.

Przytoczone tu przykłady 1–6 potwierdzają złożoność procedury nanoszenia anotacji semantycznej, unaoczniają niebezpieczeństwa czyhające na anotatora, który powodowany stanem formalnym polszczyzny błędnie będzie interpretować znaczenia kwantyfikacyjne, myląc – podane tu jako przykład – chociażby znaczenia egzystencjalne i ogólne.

4.6. Anotacja w korpusie polsko-bułgarsko-rosyjskim.

Przykład:

Svetoslav Minkov <i>Król Bezsenek</i>	Светослав Минков <i>Цар Безсънко</i>	Светослав Минков <i>Царь Бессон</i>
$\langle (\exists X)P(X)(state\_2) \rangle$ Był sobie \> $\langle ?(\exists x)(P(x)) \rangle$ krasnoludek \> z długą białą brodą, w czerwonej czapce.	$\langle (\exists X)P(X)(state\_2) \rangle$ Имаше някога \> $\langle (\exists x)P(x) \rangle$ едно джудже \> с дълга бяла брадица и алена шапчица.	$\langle (\exists X)P(X)(state\_2) \rangle$ Жил некогда \> на свете $\langle ?(\exists x)P(x) \rangle$ гном \> с длинной белой бородой.
$\langle ?(ix)P(x) \rangle$ Paletko \> $\langle (\forall X)P(X)(state\_1) \rangle$ miał \> niebieskie jak niebo, a $\langle (ix)P(x) \rangle$ na czubkach jego butów \> $\langle (\exists X)P(X)(state\_2) \rangle$ pobrzękiwały \> wesoło dwa srebrne dzwoneczki.	$\langle (ix)P(x) \rangle$ Палтенцето му \> $\langle (\forall X)P(X)(state\_1) \rangle$ беше синьо \> като небето, а $\langle (ix)P(x) \rangle$ на върховете на обувката му \> $\langle (\exists X)P(X)(state\_2) \rangle$ дрънкаха \> весело две сребърни звънчета.	$\langle (\forall X)P(X)(state\_1) \rangle$ Носил \> $\langle (ix)P(x) \rangle$ он \> алюю шапочку и синее, как небо, $\langle ?(ix)P(x) \rangle$ пальтишко \>. А $\langle (ix)P(x) \rangle$ на башмачках у него \> весело $\langle (\exists X)P(X)(state\_2) \rangle$ позвани- вали \> два серебряных колокольчика.

Svetoslav Minkov <i>Król Bezsenek</i>	Светослав Минков <i>Цар Безсънко</i>	Светослав Минков <i>Царь Бессон</i>
<(ix)P(x) Jego dom \> <(ix)P(X)_(state_1) był schowany \> jak grzyb w gęstej trawie, wśród wysokich drzew starego lasu.	<(ix)P(x) Къщицата на това джудже \> <(ix)P(X)_(state_1) стоеше \> сгушена като гъба между дърветата на една стара гора.	<(ix)P(x) Избушка гнома \> <(ix)P(X)_(event_1) при-таилась \>, словно грибок, под деревьями в старом лесу.

4.7. Anotacja w korpusie polsko-litewskim. Przykład⁷:

Antoine de Saint-Exupéry <i>Mały Księżę</i>	Antuanas de Sent-Egziuperi <i>Mažasis princas</i>
<(ix)P(X)_(state_1) Przepraszam \> <(∀x)P(x) wszystkie dzieci \> za poświęcenie <(ix)P(x) tej książki \> <?(∃x)P(x) dorosłemu \>.	<(ix)P(X)_(state_1) Atsiprašau \> <?(∀x)P(x) vaikus \>, kad <(ix)P(x) šią knygą \> <?(∃X)P(X)_(event_1) paskyriau \> <?(∃x)P(x) suaugusiam \>.
<(∀X)P(X)_(state_1) Mam <(∃x)P(x) ważne <(ix)P(x) ku temu powody \> \>: <(ix)P(x) ten dorosły \> <(∀X)P(X)_(state_1) jest <(ix)P(x) moim najlepszym przyjacielem na świecie \> \>.	<(∀X)P(X)_(state_1) Turiu <(∃x)P(x) rimtą pasiteisinimą \> \>: <(ix)P(x) tas žmogus \> <(∀X)P(X)_(state_1) yra <(ix)P(x) geriausias pasaulyje mano draugas \> \>.
<(ix)P(x) Drugi powód \>: <(∀X)P(X)_(state_2) <(ix)P(x) ten dorosły \> potrafi zrozumieć <(∀x)P(x) wszystko \> \>, nawet <czysta predykcja książki dla dzieci \>.	<(∀X)P(X)_(state_1) Turiu dar <(ix)P(x) kitą pasiteisinimą \> \>: <(∀X)P(X)_(state_2) <(ix)P(x) tas žmogus \> gali <(∀x)P(x) viską \> suprasti \>, net <czysta predykcja knygąs vaikams \>.

5. Podsumowanie

Przedstawiona w szczegółach anotacja semantyczna dotycząca kwantyfikacji zakresowej i zastosowana po raz pierwszy w wielojęzycznych korpusach równoległych z językiem polskim tworzonych przez Zespół Lingwistyki Korpusowej i Semantyki (IS PAN) jest odpowiedzią na postawiony strategiczny cel naukowy Clarin ERIC i Clarin-PL, jakim jest wspieranie badań

⁷ W tym przykładzie kolorami rozróżniono opisy odnoszące się do fraz nominalnych (kolorem niebieskim) i do fraz werbalnych (predykatów) (kolorem zielonym).

humanistycznych w wielokulturowej i wielojęzycznej Europie. Dla polskich badaczy ważna jest budowa pomostu między językiem polskim i polskimi technologiami językowymi a innymi językami europejskimi i technologiami językowymi opracowanymi dla nich. Dotychczas w nauce polskiej największy nacisk był kładziony na powiązania polsko-angielskie. Dlatego Clarin-PL podjął się opracowania pierwszych jak dotąd wielojęzycznych korpusów języka polskiego w zestawieniu z innymi językami słowiańskimi oraz z językiem litewskim: *Korpus równoległy polsko-bułgarsko-rosyjski* i *Korpus równoległy polsko-litewski*. To właśnie te korpusy przełamują dotychczasowe „kanony” i udostępniają nauce powiązane wielojęzyczne zasoby z językiem polskim – w pierwszym etapie ograniczone do języków-przedstawicieli trzech grup słowiańskich oraz języka litewskiego.

Zaprezentowana tu anotacja semantyczna, jako novum w lingwistyce korpusowej, jest póki co nanoszona ręcznie. Identyfikacja poszczególnych wartości kwantyfikacji zakresowej w zdaniu oraz przedstawiane w artykule próby jej zapisu są poparte wieloletnimi badaniami międzynarodowego zespołu lingwistów i matematyków-informatyków, opracowujących zagadnienie kwantyfikacji imion, czasu i aspektu w językach naturalnych. Cechą charakterystyczną i znamioną tej anotacji semantycznej jest zastosowanie do opisu poszczególnych znaczeń kwantyfikatorów (ogólnego, egzystencjalnego i operatora jota). W odróżnieniu od innych znanych w literaturze przedmiotu anotacji semantycznych ta nie odnosi się do semantyki leksykalnej, FrameNetu czy WordNetu. Bazuje na wypracowanej przez Zespół Lingwistyki Korpusowej i Semantyki IS PAN koncepcji teoretycznych badań konfrontatywnych z językiem-pośrednikiem, w której wyraźnie rozróżnia się płaszczyznę znaczeniową i płaszczyznę formalną.

Podczas ręcznego nanoszenia anotacji zauważono pewne możliwości automatyzacji tegoż procesu. W obu językach stwierdza się bowiem istnienie jednoznacznych wykładników kwantyfikacji zakresowej, które niezależnie od kontekstu zawsze wyrażają jedną i tę samą kwantyfikację. Szczególną rolę w procesie automatyzacji nanoszenia anotacji semantycznej autorzy upatrują w przejrzystej strukturze języka litewskiego, dla którego stwierdza się wiele jednoznacznych powiązań między formą a znaczeniem, oraz w późniejszym rzutowaniu ustalonych znaczeń kwantyfikacyjnych form litewskich na polskie odpowiedniki⁸. Dzięki temu zabiegowi wartości opisywanej anotacji seman-

---

⁸ Przedstawienie zagadnienia rzutowania anotacji w wielojęzycznych korpusach jest planowane na seminarium Clarin-PL na kwiecień 2016 roku (Roman Roszko, *Korpusy wie-*

tycznej zostaną powiązane z planem formalnym, co jest jednym z warunków gwarantujących stosowanie i rozwój tej anotacji w korpusach wielojęzycznych.

Bibliografia

- Ajdukiewicz, K. (1965). *Logika pragmatyczna*. Warszawa: PWN.
- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. W: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (T. 1, ss. 86–90). Stroudsburg, PA: Association for Computational Linguistics. <http://dx.doi.org/10.3115/980845.980860>
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2), 159–219. <http://dx.doi.org/10.1007/BF00350139>
- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: Bradford Books; MIT Press.
- Bellert, I. (1971). On the use of linguistic quantifying operators. *Poetics*, 1(2), 71–86. [http://dx.doi.org/10.1016/0304-422X\(71\)90011-8](http://dx.doi.org/10.1016/0304-422X(71)90011-8)
- Bogusławski, A. (2003). *Aspekt i negacja*. Warszawa: Katedra Lingwistyki Formalnej UW.
- Bond, F., Morgado da Costa, L., & Lê, T. A. (2015). IMI – A Multilingual Semantic Annotation Environment. W: *Proceedings of ACL-IJCNLP 2015 System Demonstrations Beijing, China, July 26–31, 2015* (ss. 7–12). <http://dx.doi.org/10.3115/v1/P15-4002>
- Cooper, R. (1996). The role of situations and generalized quantifiers. W: S. Lappin (Red.), *The handbook of contemporary semantic theory*. Oxford: Blackwell Reference.
- Desclés, J. P. (1999). Quantification, types, preuves et logique combinatoire. W: *Études cognitives 3. Quantification, Temps, Aspects* (ss. 13–82). Varsovie: SOW.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Phil. und phil. NF Kritik*, 100, 25–50.
- Grzegorzczak, R. (1972). Wykładniki kwantyfikacji w polskim zdaniu. W: *Z polskich studiów slawistycznych: Seria 4: Prace na VII Międzynarodowy Kongres Slawistów w Warszawie, 1973* (T. 1, *Językoznawstwo*, ss. 13–19). Warszawa: PWN.

---

lojęzyczne. Rzutowanie anotacji z jednego języka na drugi: Abstract: Zachowawczy charakter języka litewskiego, przejrzystość struktury formalnej oraz proste (nieuwikłane) odniesienie płaszczyzn formalnej i znaczeniowej pozwalają na wzbogacenie opisu innych, odniesionych do języka litewskiego zasobów językowych, o wartości przynależne płaszczyźnie znaczeniowej. Omawiana projekcja anotowanych wartości z języka litewskiego na polski lub inny język pozwala nie tylko automatycznie ujednoznaczyć wieloznaczne formy innego języka, lecz także dodać te wartości, które w innych językach nie są kodowane.).

- Grzegorzczak, R. (1976). Aktualizacja wypowiedzi poprzez kwantyfikację argumentów i predykatu. *Otázky Slovanské syntaxe*, 4(1), 195–200.
- Impact of European integration on Slavic languages: Trends in translation and multilingual communication*. (2009). (D. Čaplovič, K.-J. Lönnroth, & P. Zsapka, Org.). Conference and Directorate-General for Translations of the European Commission. Bratislava.
- Karolak, S. (2008). *Gramatyka konfrontatywna bułgarsko-polska* (T. 8, *Semantyczna kategoria aspektu*). (V. Koseska-Toszeza & J. Penčev, Red.). Warszawa: SOW.
- Kingsbury, P., Palmer, M., & Marcus, M. (2002). Adding semantic annotation to the Penn TreeBank. W: *Proceedings of the Human Language Technology Conference (HLT 2002)*. Pobrano 1 kwietnia 2016 z http://faculty.washington.edu/fxia/courses/LING571/PropBank_HLT2002.pdf
- Kisiel, A., Koseska-Toszeza, V., Kotsyba, N., Satoła-Staśkowiak, J., & Sosnowski, W. (2016). Polish-Bulgarian-Russian Parallel Corpus. CLARIN-PL digital repository. Pobrano 4 września 2016 z <http://hdl.handle.net/11321/308>
- Koseska-Toszeza, V. (1982). *Semantyczne aspekty kategorii określoności/nieokreśloności (na materiale z języka bułgarskiego, polskiego i rosyjskiego)*. Wrocław: Ossolineum.
- Koseska-Toszeza, V. (1995). O aspekcie i czasie w ujęciu kognitywnym. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 51, 25–31.
- Koseska-Toszeza, V. (1997). Квантификация и вид. W: *Семантика и структура славянского вида* (T. 2, ss. 141–147). Kraków: Wydawnictwo Naukowe Wyższej Szkoły Pedagogicznej.
- Koseska-Toszeza, V. (2006). *Gramatyka konfrontatywna bułgarsko-polska* (T. 7, *Semantyczna kategoria czasu*). (V. Koseska-Toszeza & J. Penčev, Red.). Warszawa: SOW.
- Koseska-Toszeza, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie „Dialog”.
- Koseska-Toszeza, V., & Mazurkiewicz, A. (1988). Net representation of sentences in natural languages. W: *Lecture Notes in Computer Science 340. Advances in Petri Nets 1988* (ss. 249–266). Berlin: Springer Verlag.
- Koseska-Toszeza, V., & Mazurkiewicz, A. (1994). Description à l'aide de réseaux de la temporalité et modalité dans la phrases dans la langue naturelle. W: *Studia kognitywne* (T. 1, ss. 89–112). Warszawa: SOW.
- Koseska-Toszeza, V., & Mazurkiewicz, A. (2010). *Time flow and tenses*. Warszawa: SOW.
- Laskowski, R. (2011). Българско-полска съпоставителна граматика, София – Варшава 1988–2011, vol. 1–9. *Български език*, 58(3), 172–176.
- Mazurkiewicz, A. (1986). Zdarzenia i stany: Elementy temporalności. W: V. Koseska-Toszeza, I. Sawicka, & J. Mindak (Eds.), *Studia gramatyczne bułgarsko-polskie* (T. 1, *Temporalność*, ss. 7–21). Wrocław: Ossolineum.
- Mazurkiewicz, A., & Koseska-Toszeza, V. (1991). Sieciowe przedstawienie temporalności i modalności w zdaniach języka naturalnego. W: V. Koseska-Toszeza, I. Sawicka, & J. Mindak (Eds.), *Studia gramatyczne bułgarsko-polskie* (T. 1, *Temporalność*, ss. 7–21). Wrocław: Ossolineum.

- Padó, S., & Lapata, M. (2009). Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, 36, 307–340.
- Petri, C. A. (1962). Fundamentals of the theory of asynchronous information flow. W: *Proc. of IFIP'62 Congress* (ss. 386–391). Amsterdam: North Holland Publ. Comp.
- Rasiowa, H. (1975). *Wstęp do matematyki współczesnej*. Warszawa: PWN.
- Reeder, F., Miller, K., Farwell, D., Helmreich, S., Hovy, E., Levin, L., ... Siddharthan, A. (2004). Semantic annotation of multilingual text corpora. W: *Proceedings of COLING, Geneva, Switzerland*. Pobrano 1 kwietnia 2016 z <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.643.5915&rep=rep1&type=pdf>
- Reichenbach, H. (1967). Analiza języka potocznego. W: J. Pelc (Red.), *Logika i język* (ss. 85–222). Warszawa: PWN.
- Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (Na tle literackich języków polskiego i litewskiego)*. Warszawa: SOW.
- Roszko, D., & Roszko, R. (2007). Функции литовского прошедшего многократного времени (русские соответствия). W: S. Polkovnikova (Red.), *Valoda, Valoda dažādu kultūru kontekstā, Zinātnisko rakstu krājums* (T. 17, ss. 200–207). Daugavpils: Daugavpils Universitātes, Akadēmiskais apgāds «Saule».
- Roszko, D., & Roszko, R. (2014). A net presentation of Lithuanian sentences containing verbal forms with the grammatical suffix -dav-. *Cognitive Studies | Études cognitives*, (14), 173–182. <http://dx.doi.org/10.11649/cs.2014.014>
- Roszko, D. & Roszko, R. (2016). Polish-Lithuanian Parallel Corpus. CLARIN-PL digital repository. Retrieved 4 September 2016 from <http://hdl.handle.net/11321/309>
- Roszko, R. (2004). *Semantyczna kategoria określoności/nieokreśloności w języku litewskim (w zestawieniu z językiem polskim)*. Warszawa: SOW.
- Russell, B. (1967). Denotowanie. Deskrypcje. W: J. Pelc (Red.) *Logika i język* (ss. 377–413). Warszawa: PWN.
- Volk, M., & Samuelsson, Y. (2007). Frame-semantic annotation on a parallel treebank. W: *Proc. of Nodalida Workshop on Building Frame Semantics Resources for Scandinavian and Baltic Languages*. Tartu. Pobrano 1 kwietnia 2016 z <http://fileadmin.cs.lth.se/nlp/frame2007/Proceedings/MVolk.pdf>
- Иванчев, С. (1971). *Проблеми на аспектиалността в славянските езици*. София: Издателство на БАН.
- Косеска-Тошева, В., & Гаргов, Г. (1990). *Българско-полска съпоставителна граматика (Т. 2, Семантичната категория определеност/неопределеност)*. София: Издателство на БАН.

Bibliography (Transliteration)

- Ajdukiewicz, K. (1965). *Logika pragmatyczna*. Warszawa: PWN.
- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (Vol. 1, pp. 86–90). Stroudsburg, PA: Association for Computational Linguistics. <http://dx.doi.org/10.3115/980845.980860>
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2), 159–219. <http://dx.doi.org/10.1007/BF00350139>
- Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: Bradford Books; MIT Press.
- Bellert, I. (1971). On the use of linguistic quantifying operators. *Poetics*, 1(2), 71–86. [http://dx.doi.org/10.1016/0304-422X\(71\)90011-8](http://dx.doi.org/10.1016/0304-422X(71)90011-8)
- Bogusławski, A. (2003). *Aspekt i negacja*. Warszawa: Katedra Lingwistyki Formalnej UW.
- Bond, F., Morgado da Costa, L., & Lê, T. A. (2015). IMI – A Multilingual Semantic Annotation Environment. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations Beijing, China, July 26–31, 2015* (pp. 7–12). <http://dx.doi.org/10.3115/v1/P15-4002>
- Cooper, R. (1996). The role of situations and generalized quantifiers. In S. Lappin (Ed.), *The handbook of contemporary semantic theory*. Oxford: Blackwell Reference.
- Desclés, J. P. (1999). Quantification, types, preuves et logique combinatoire. In *Études cognitives 3. Quantification, Temps, Aspects* (pp. 13–82). Varsovie: SOW.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Phil. und phil. NF Kritik*, 100, 25–50.
- Grzegorzczak, R. (1972). Wykładowcy kwantyfikacji w polskim zdaniu. In *Z polskich studiów slawistycznych: Seria 4: Prace na VII Międzynarodowy Kongres Slawistów w Warszawie, 1973* (Vol. 1, *Językoznawstwo*, pp. 13–19). Warszawa: PWN.
- Grzegorzczak, R. (1976). Aktualizacja wypowiedzi poprzez kwantyfikację argumentów i predykatu. *Otázky Slovanské syntaxe*, 4(1), 195–200.
- Impact of European integration on Slavic languages: Trends in translation and multilingual communication*. (2009). (D. Čaplovič, K.-J. Lönnroth, & P. Zsapka, Orgs.). Conference and Directorate-General for Translations of the European Commission. Bratislava.
- Ivanchev, S. (1971). *Problemi na aspektualnostta v slavianskite ezitsi*. Sofia: Izdatelstvo na BAN.
- Karolak, S. (2008). *Gramatyka konfrontatywna bułgarsko-polska* (T. 8, *Semantyczna kategoria aspektu*). (V. Koseska-Toszewa & J. Penčev, Eds.). Warszawa: SOW.
- Kingsbury, P., Palmer, M., & Marcus, M. (2002). Adding semantic annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference (HLT 2002)*. Retrieved 1 April 2016 from http://faculty.washington.edu/fxia/courses/LING571/PropBank_HLT2002.pdf
- Kisiel, A., Koseska-Toszewa, V., Kotsyba, N., Satoła-Staśkowiak, J., & Sosnowski, W. (2016). Polish-Bulgarian-Russian Parallel Corpus, CLARIN-PL digital repository. Retrieved 4 September 2016 from <http://hdl.handle.net/11321/308>

- Koseska-Tosheva, V., & Gargov, G. (1990). *Bŭlgarsko-polska sŭpostavitelna gramatika* (Vol. 2, *Semantichnata kategoriia opredelenost/neopredelenost*). Sofia: Izdatelstvo na BAN.
- Koseska-Toszewa, V. (1982). *Semantyczne aspekty kategorii określoności/nieokreśloności (na materiale z języka bułgarskiego, polskiego i rosyjskiego)*. Wrocław: Ossolineum.
- Koseska-Toszewa, V. (1995). O aspekcie i czasie w ujęciu kognitywnym. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 51, 25–31.
- Koseska-Toszewa, V. (1997). Kvantifikatsiia i vid. In *Semantika i struktura slavianskogo vida* (Vol. 2, pp. 141–147). Kraków: Wydawnictwo Naukowe Wyższej Szkoły Pedagogicznej.
- Koseska-Toszewa, V. (2006). *Gramatyka konfrontatywna bułgarsko-polska* (Vol. 7, *Semantyczna kategoria czasu*). (V. Koseska-Toszewa & J. Penčev, Eds.). Warszawa: SOW.
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie „Dialog”.
- Koseska-Toszewa, V., & Mazurkiewicz, A. (1988). Net representation of sentences in natural languages. W: *Lecture Notes in Computer Science 340. Advances in Petri Nets 1988* (pp. 249–266). Berlin: Springer Verlag.
- Koseska-Toszewa, V., & Mazurkiewicz, A. (1994). Description à l'aide de réseaux de la temporalité et modalité dans la phrases dans la langue naturelle. In *Studia kognitywne* (T. 1, ss. 89–112). Warszawa: SOW.
- Koseska-Toszewa, V., & Mazurkiewicz, A. (2010). *Time flow and tenses*. Warszawa: SOW.
- Laskowski, R. (2011). Bŭlgarsko-polska sŭpostavitelna gramatika, Sofia – Varshava 1988–2011, vol. 1–9. *Bŭlgarski ezik*, 58(3), 172–176.
- Mazurkiewicz, A. (1986). Zdarzenia i stany: Elementy temporalności. In V. Koseska-Toszewa, I. Sawicka, & J. Mindak (Eds.), *Studia gramatyczne bułgarsko-polskie* (Vol. 1, *Temporalność*, pp. 7–21). Wrocław: Ossolineum.
- Mazurkiewicz, A., & Koseska-Toszewa, V. (1991). Sieciowe przedstawienie temporalności i modalności w zdaniach języka naturalnego. In V. Koseska-Toszewa, I. Sawicka, & J. Mindak (Eds.), *Studia gramatyczne bułgarsko-polskie* (Vol. 1, *Temporalność*, pp. 7–21). Wrocław: Ossolineum.
- Padó, S., & Lapata, M. (2009). Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, 36, 307–340.
- Petri, C. A. (1962). Fundamentals of the theory of asynchronous information flow. In *Proc. of IFIP'62 Congress* (pp. 386–391). Amsterdam: North Holland Publ. Comp.
- Rasiowa, H. (1975). *Wstęp do matematyki współczesnej*. Warszawa: PWN.
- Reeder, F., Miller, K., Farwell, D., Helmreich, S., Hovy, E., Levin, L., ... Siddharthan, A. (2004). Semantic annotation of multilingual text corpora. In *Proceedings of COLING, Geneva, Switzerland*. Retrieved 1 April 2016 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.643.5915&rep=rep1&type=pdf>
- Reichenbach, H. (1967). Analiza języka potocznego. In J. Pelc (Ed.), *Logika i język* (pp. 85–222). Warszawa: PWN.

- Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (Na tle literackich języków polskiego i litewskiego)*. Warszawa: SOW.
- Roszko, D., & Roszko, R. (2007). Funktsii litovskogo proshedshego mnogokratnogo vremeni (russkie sootvetstviia). In S. Polkovnikova (Ed.), *Valoda, Valoda dažādu kultūru kontekstā, Zinātnisko rakstu krājums* (Vol. 17, pp. 200–207). Daugavpils: Daugavpils Universitātes, Akadēmiskais apgāds „Saulė”.
- Roszko, D., & Roszko, R. (2014). A net presentation of Lithuanian sentences containing verbal forms with the grammatical suffix -dav-. *Cognitive Studies | Études cognitives*, (14), 173–182. <http://dx.doi.org/10.11649/cs.2014.014>
- Roszko, D. & Roszko, R. (2016). Polish-Lithuanian Parallel Corpus, CLARIN-PL digital repository. <http://hdl.handle.net/11321/309>
- Roszko, R. (2004). *Semantyczna kategoria określoności/nieokreśloności w języku litewskim (w zestawieniu z językiem polskim)*. Warszawa: SOW.
- Russell, B. (1967). Denotowanie. Deskrypcje. In J. Pelc (Ed.) *Logika i język* (pp. 377–413). Warszawa: PWN.
- Volk, M., & Samuelsson, Y. (2007). Frame-semantic annotation on a parallel treebank. In *Proc. of Nodalida Workshop on Building Frame Semantics Resources for Scandinavian and Baltic Languages*. Tartu. Retrieved 1 April 2016 from <http://fileadmin.cs.lth.se/nlp/frame2007/Proceedings/MVolk.pdf>

Slavic languages and the Lithuanian language in the Clarin-PL parallel corpora

Summary

The Clarin Eric and Clarin-PL strategic scientific purpose is to support humanistic research in a multicultural and multilingual Europe. Polish researchers put the emphasis on building a bridge between the Polish language and Polish linguistic technologies and other European languages and their linguistic technologies. So far, the Polish scientific community has mainly focused on Polish-English connections. Clarin-PL has been developing the first and only multilingual corpora of the Polish language in conjunction with other Slavic languages and the Lithuanian language: the *Polish-Bulgarian-Russian Parallel Corpus* and the *Polish-Lithuanian Parallel Corpus*. The parallel corpora created by the ISS PAS Corpus

Linguistics and Semantics Team break through the existing “canons” and allow scientists access to interlinked multilingual language resources – in the first phase limited to the languages of the three Slavic groups and the Lithuanian language. In the article, the authors present very detailed information on their original system of the semantic annotation of scope quantification in multilingual parallel corpora, hitherto unused in the subject literature. Due to the system’s originality, the semantic annotation is carried out manually. Identification of particular values of scope quantification in a sentence and the hereby presented attempts of its recording are supported by long-term research conducted by an international team of linguists and computer scientists / mathematicians developing the issue of quantification of names, time and aspect in natural languages.

Języki słowiańskie i litewski w korpusach równoległych Clarin-PL

Streszczenie

Strategicznym celem naukowym Clarin ERIC i Clarin-PL jest wspieranie badań humanistycznych w wielokulturowej i wielojęzycznej Europie. Dla polskich badaczy ważna jest budowa pomostu między językiem polskim, polskimi technologiami językowymi a innymi językami europejskimi i na ich rzecz opracowanymi technologiami językowymi. Dotychczas w nauce polskiej największy nacisk był kładziony na powiązania polsko-angielskie. Clarin-PL opracowuje zatem pierwsze jak dotąd wielojęzyczne korpusy języka polskiego w zestawieniu z innymi językami słowiańskimi oraz z językiem litewskim: *Korpus równoległy polsko-bułgarsko-rosyjski* i *Korpus równoległy polsko-litewski*. Tworzone przez Zespół Lingwistyki Korpusowej i Semantyki (IS PAN) korpusy równoległe przełamują dotychczasowe „kanony” i udostępniają nauce powiązane wielojęzyczne zasoby – w pierwszym etapie ograniczone do języków trzech grup słowiańskich oraz języka litewskiego. W artykule autorzy przedstawiają bardzo szczegółową informację o zastosowanej po raz pierwszy w literaturze przedmiotu anotacji semantycznej dotyczącej kwantyfikacji zakresowej w wielojęzycznych korpusach równoległych. Z powodu swojego rozległego zakresu

i nowatorstwa ta anotacja semantyczna jest nanoszona ręcznie. Identyfikacja poszczególnych wartości kwantyfikacji zakresowej w zdaniu oraz przedstawiane tu próby jej zapisu są poparte wieloletnimi badaniami międzynarodowego zespołu lingwistów i matematyków-informatyków opracowujących zagadnienie kwantyfikacji imion, czasu i aspektu w językach naturalnych.

Keywords: multilingual parallel corpora; semantic annotation; scope quantification

Słowa kluczowe: wielojęzyczne korpusy równoległe; anotacja semantyczna; kwalifikacja zakresowa

---

(1) Violetta Koseska-Toszewa, Professor Emerita, Institute of Slavic Studies, Polish Academy of Sciences, Warsaw
Correspondence: amaz@inetia.pl

(2) Roman Roszko, Institute of Slavic Studies, Polish Academy of Sciences, Warsaw
Correspondence: roman.roszko@ispan.waw.pl

Authors contribution: article idea (V.K.-T. and R.R.); description of the theoretical framework, Polish, Bulgarian and Russian texts annotation (V.K.-T.); description of the state of research, material analysis, Polish and Lithuanian texts annotation (R.R.).

This work was funded by the Ministry of Science and Higher Education within the programme CLARIN ERIC for 2015–2016 (<http://clarin.eu/>).

Competing interests: The authors have declared they have no competing interests.