

Marek Łaziński
Uniwersytet Warszawski

Magdalena Kuratczyk
Uniwersytet Warszawski

Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego

The University of Warsaw Polish-Russian Parallel Corpus

Abstract

The *Polish-Russian Parallel Corpus* has been developed at the University of Warsaw (the Faculty of Polish Studies and the Institute of Russian Studies) in co-operation with the *National Corpus of Polish* and the *Russian National Corpus*.

The corpus consists of Russian and Polish literary classics (90%), nonfiction books, legal texts (5%), religious texts (i.e. Bible translations; 4%) and contemporary press articles (1%). Great Russian realistic novels of the 19th century, together with modern Russian books which are most popular in Poland, made up a significant part of the corpus. We have also taken into account these works of Polish literature that are the most widely known in Russia. Looking for *loci communes* in the Russian and Polish cultures was an important, extra-linguistic aspect of the corpus project.

Unfortunately, the novels by Dostoevsky or Tolstoy were translated into Polish only in the thirties and the copyright protection for the translations – 70 years since the author's death – is still in force. Some of the translators' heirs did not grant their permission to include the texts in the corpus.

The annotation and search possibilities in the corpus result from co-operation with the national corpora. However, not all levels of annotation applied in the source corpora will be used in the parallel corpus. Two national corpora differ according to grammatical disambiguation of annotated word forms. In the *National Corpus of Polish* all texts are disambiguated, in the *Russian National Corpus* only some of them have undergone this procedure.

The search interface is based on the user-friendly interface of the *Russian National Corpus*. It allows formulating lexical and grammatical queries using the tags present in the tag sets of the two national corpora and is easy for users of both national corpora.

In the second part of the paper some practical applications of the corpus in the linguistic research, translation practice and foreign language teaching have been shown. The first case is a Russian translation

strategy of the Polish unspecified numeral *kilkanaście* ‘over a dozen’, the second – a Polish translation strategy of the Russian adjectives *russkij* and *rossijskij* ‘Russian’.

Słowa kluczowe: korpus równoległy, język polski, język rosyjski, zastosowania praktyczne, leksykografia dwujęzyczna, analiza kontrastywna

Keywords: parallel corpus, Polish, Russian, practical applications, bilingual lexicography, contrastive analysis

1. Wprowadzenie

Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego (www.pol-ros.polon.uw.edu.pl) to reprezentatywny (ze względu na zróżnicowanie tekstów), anotowany morfosyntaktycznie i dezambiguowany (ujednoznaczony morfologicznie) korpus równoległy opatrzony informacjami bibliograficznymi. Tak jak inne korpusy równoległe, i ten może być praktyczną pomocą dla tłumaczy, lingwistów, leksykoграфów, badaczy kultury i literatury.

2. Geneza projektu

Korpus był pierwszym i do roku 2014 jedynym dostępnym publicznie oraz bezpłatnie korpusem równoległym dwóch języków. Obecnie jest więcej takich korpusów, m.in. polsko-angielski zbiór Paralela na stronie clarin.eu/pl. Pomysł stworzenia korpusu powstał w roku 2009 w trakcie prac nad Narodowym Korpusem Języka Polskiego (nkjp.pl). Zwrócili się z nim do polskich środowisk korpusowych lingwiści z Uniwersytetu Pedagogicznego w Ufie: Elena Slobodyan i Boris Orekhov w porozumieniu z innymi środowiskami korpusowymi w Rosji.

Wkrótce po polskiej stronie złożono wniosek do MNiSW¹, który uzyskał grant na realizację projektu NN104056638. Kierownikiem projektu polskiego został Marek Łaziński z Instytutu Języka Polskiego UW, ze strony Instytutu Rusycystyki projekt koordynowała Magdalena Kuratczyk, od początku w projekt była zaangażowana Natalia Godlewska, doktorantka w Instytucie Języka Polskiego UW. Ze strony rosyjskiej istotnej, a od roku 2012 decydującej pomocy w powstaniu korpusu udzielał zespół Narodowego Korpusu Języka Rosyjskiego z Dmitrijem Sičinawą i Svetlaną Minlos. Po polskiej stronie pomocą służył Instytut Podstaw Informatyki, który wcześniej był głównym wykonawcą Narodowego Korpusu Języka Polskiego. Istotna część tekstów literackich Korpusu Polsko-Rosyjskiego, była także w zestawie NKJP. W naszym projekcie wykorzystano

¹ W roku 2011 agencją wykonawczą dla opisywanego projektu, tak jak dla większości grantów naukowych, stało się Narodowe Centrum Nauki.

też tager IPIPAN TAKIPI, stosowany w pierwotnej wersji NKJP (obecna wersja NKJP jest znakowana tagerem Pantera).

Założenia przyjęte na początku projektu zostały opisane w artykule (Łaziński et al. 2012), a zakończony projekt – w artykule (Łaziński, Kuratczyk, 2015)².

3. Struktura korpusu

Korpus liczy w całości około 30 milionów segmentów liczonych według tagera TAKIPI i większości polskich korpusów za segmenty uważa się nie tylko formy wyrazowe w rozumieniu tradycyjnej gramatyki, lecz także aglutynacyjne morfemy osobowo-liczbowe czasownika w (*-em, -eś, -śmy, -ście*, każdy morfem *się* lub *by* niezależnie od ich statusu słowotwórczego lub fleksyjnego).

Połowa tekstów jest w wersji polskiej, połowa w rosyjskiej, z nieznaczną przewagą liczby słów rosyjskich. Jednak liczba polskich tekstów oryginalnych jest wyższa niż rosyjskich. Osiągnięcie planowanej na początku idealnej równowagi okazało się w toku zbierania tekstów niemożliwe. Ostatecznie mamy 50 procent oryginałów polskich, 33 procent oryginałów rosyjskich i pozostałe 17 procent oryginałów z trzeciego języka (w tym duży udział tekstów biblijnych, a także przekłady traktatów międzynarodowych).

Projekt korpusu zakładał włączenie znaczącego udziału literatury pięknej, w tym klasyki dziewiętnastowiecznej, która nie podlega już ochronie majątkowych praw autorskich. Beletrystyka stanowi zresztą zdecydowaną większość licznych korpusów równoległych, np. ParaSol. W naszym korpusie teksty literatury pięknej stanowią 90 procent. Staraliśmy się jednak włączyć doń teksty prasowe: przekłady tekstów rosyjskich pochodzące z tygodnika „Forum” oraz przekłady tekstów polskich z portalu inosmi.ru. Choć te teksty to tylko jeden procent korpusu, są one niezastąpionym źródłem do badań najnowszego słownictwa wspólnostylowego i potocznego. Nasz korpus jest pod tym względem wyjątkiem na tle innych znacznie większych i bardziej znanych. Np. korpus równoległy InterCorp Czeskiego Korpusu Narodowego zawiera teksty prasowe z portali PressEurope (presseurope.eu) i Syndicate (project-syndicate.org), jednak ich tematyka jest ograniczona do polityki Wspólnoty Europejskiej i spraw tak samo dobrze znanych w całej Europie (zresztą w trakcie tworzenia korpusu projekt PressEurope nie miał wersji rosyjskojęzycznej, Syndicate – polskojęzycznej).

² Artykuł ten, napisany po rosyjsku pokrywa się treściowo w dużej części z niniejszym tekstem, który jest najpełniejszym opisem korpusu w języku polskim.

Korpus zawiera aż 4% tekstów religijnych, w tym najczęściej cytowane w obu krajach przekłady Biblii: Biblia Tysiąclecia (1965-1971) oraz przekład synodalny (1867-1876). Choć w sieci jest wiele serwisów proponujących przekłady Biblii w wielu językach, np. biblia.net.pl, biblia.info.pl, bibleonline.ru, to w żadnym z nich nie znajdujemy jednocześnie tych dwóch przekładów, do wyjątków należą serwisy zawierające jakiegokolwiek przekłady polskie i rosyjskie jednocześnie (biblegateway.com).

Pozostałe 5 procent korpusu stanowią teksty prawne, literatura popularno-naukowa, wspomnieniowa i szerzej non-fiction.

Okolo 66 procent tekstów dostępnych jest publicznie bez logowania. Pozostałe teksty, w sprawie których korespondencja z posiadaczami praw autorskich jeszcze się nie zakończyła, dostępne są lokalnie dla współpracowników projektu na Uniwersytecie Warszawskim³. W ramach współpracy z Narodowym Korpusem Języka Rosyjskiego w gromadzeniu części tekstów dla naszego korpusu oraz dla powstającego w Moskwie rosyjsko-polskiego korpusu porównawczego zostawiliśmy ochronę praw autorskich do tekstów rosyjskich (oryginałów i przekładów z polskiego) partnerom rosyjskim.

Warto zauważyć, że problemy z uzyskaniem zgody posiadaczy praw autorskich dotyczyły nie tylko tekstów współczesnych, lecz nawet rosyjskiej klasyki XIX wieku. Powieści Dostojewskiego czy Tołstoja długo nie były tłumaczone na język polski, ponieważ, kiedy się ukazywały, urzędowym językiem nad Wisłą był rosyjski. Były przekładane już w wolnej Polsce, w latach dwudziestych i trzydziestych lub po wojnie, przez tłumaczy w większości młodych, którzy zmarli w latach siedemdziesiątych i później, więc ochrona ich praw majątkowych będzie trwać jeszcze długo (zgodnie z obowiązującą Ustawą o prawie autorskim – 70 lat po śmierci twórcy).

4. Aspekt kulturowy budowy korpusu

Nasz projekt ma ambicje nie tylko dydaktyczne i naukowe, lecz i kulturowe. Staraliśmy się zgromadzić teksty autorów rosyjskich szczególnie popularnych w Polsce i ważnych dla polskiego obrazu Rosji i teksty polskich autorów popularnych w Rosji. Nie przypadkiem zaczęliśmy zbierać teksty rosyjskie Aleksandra Sołżenicyna w przekładzie Jerzego Pomianowskiego (on też – jako pierwszy spośród tłumaczy, do których się zwróciliśmy – wydał zgodę na wykorzystanie

³ Przy włączaniu tekstów do korpusu przyjęliśmy najbardziej restrykcyjną interpretację Ustawy o prawie autorskim. Zdaniem części prawników wykorzystanie dowolnych utworów nie wykraczające poza pokazywanie pojedynczych zdań lub akapitów w ogóle nie wymaga zezwolenia, jednak z drugiej strony, włączenie tekstu do korpusu w celu udostępniania cytatów wiąże się z obróbką całości utworu.

w korpusie dokonanego przez siebie przekładu)⁴. Spośród współczesnych pisarzy rosyjskich nie można było zapomnieć o Wiktorze Jerofiejewie, aktywnym uczestniku debat o stosunkach polsko-rosyjskich⁵. Przybliżeniu kultury rosyjskiej w Polsce mają służyć też kilkudziesięciu opisy bibliograficzne rosyjskich tekstów literackich daleko wykraczające poza standardowe dane metryczek korpusów. Opisy przygotowali literaturoznawcy z Instytutu Rusycystyki UW.

The screenshot shows the main interface of the 'Polsko-rosyjski i rosyjsko-polski korpus równoległy' website. The header includes the University of Warsaw logo and the title. The left sidebar contains navigation links: 'Strona główna', 'Struktura korpusu', 'Ochrona prawna tekstów', 'Zespół i instytucje współpracujące', 'Podziękowania', and 'Inne korpusy'. Below this are sections for 'Wyszukiwarki' (Wyszukiwanie słów, Wyszukiwanie morfologiczne, Jak szukać) and 'Korpus rozszerzony' (Dostęp rozszerzony włączony, Wyloguj się). The main content area is titled 'Wyszukiwanie słów' and features several search filters: 'Fakt', 'Fikcja (do 1945)', 'Fikcja (po 1945)', 'Teksty prasowe', 'Teksty prawne', 'Teksty religijne', 'Oryginały polskie', 'Oryginały rosyjskie', and 'Teksty obce'. There are two search input fields, each with a 'Słowo:' label and a text input box. Below each input field is a 'Limit wyników' dropdown menu (set to 50 and 10 respectively) and a 'Wyszukaj' button, with a 'Wyczyść' link below. A 'Wyszukiwanie łączne:' section at the bottom has two input fields with flags and a 'Wyszukaj' button, also with a 'Wyczyść' link. The footer contains copyright information: 'Copyright © 2013 Uniwersytet Warszawski' and 'Praca naukowa finansowana ze środków na naukę w latach 2010-2012 - projekt NN 104056638'.

Rysunek 1. Strona główna Korpusu Polsko-Rosyjskiego Uniwersytetu Warszawskiego (www.pol-ros.polon.uw.edu.pl)

Polscy uczestnicy projektu mogli także obejrzyć własną tradycję w zwierciadle kultury rosyjskiej. Pierwszym tekstem, o który poprosili rosyjscy partnerzy na początku projektu, był wolny od majątkowych praw autorskich *Faraon* Bolesława Prusa. Z naszego punktu widzenia *Faraon* na pewno nie jest najważniejszą powieścią dziewiętnastowieczną. Powieść dawno znikła z listy szkolnych lektur obowiązkowych, czytają ją właściwie tylko studenci polonistyki, a większość Polaków zna jedynie film Jerzego Kawalerowicza. Wybór *Faraona* można

4 Spośród dzieł A. Solżenicyna w naszym korpusie znalazły się: *Archipelag GULAG* (3 tomy) i *Krąg Pierwszy* (2 tomy) w przekładzie J. Pomianowskiego oraz *Zagroda Matryony* w tłumaczeniu A. Wołodźki.

5 Wśród tekstów W. Jerofiejewa przetłumaczonych na język polski w korpusie uwzględnione zostały: *Rosyjska apokalipsa*, *Bóg X*, *Mężczyźni* i *Życie z idiotą*.

oczywiście tłumaczyć osobistym gustem rosyjskich partnerów polonistów, ale bardziej prawdopodobną przyczyną jest zadziwiająca popularność w Rosji tej powieści celnie opisującej uniwersalne mechanizmy władzy⁶. Wyszukiwarka Google znajduje w rosyjskojęzycznym Internecie 23000 stron z *Faraonem* Prusa i tylko 2000 z *Lalką* tego samego autora. Polskojęzyczne Google znajduje 173000 stron z *Lalką* i 51000 z *Faraonem*.

5. Tagowanie i dezambiguacja tekstu oraz program wyszukujący

Teksty korpusu przechowywane są w relacyjnej bazie danych. Tabele bazy zawierają informacje o słowach, zdaniach i utworach, do których słowa należą, oraz o odpowiedności zdań między językami w wyrównanym tekście. Język zapytań SQL umożliwia pytanie o konkretne formy tekstowe, leksemy lub kategorie gramatyczne. Każde słowo lub jednostka, o którą pytamy, jest identyfikowana w zdaniu i system znajduje odpowiednie zdanie w przekładzie.

Do wyrównywania tekstów (alignment) korzystano z programu ABBY Aligner. Polskie teksty tagowano tagerem TAKIPI, rosyjskie – tagerem Mystem.

Okienkowy interfejs wyszukiwania przypomina interfejs Narodowego Korpusu Języka Rosyjskiego (ruscorporu.ru). Do wyboru mamy wyszukiwanie słów tekstowych oraz wyszukiwanie morfologiczne. Interfejs wyszukiwania morfologicznego, w którym wystarczy wybrać odpowiednią kategorię gramatyczną z listy, jest łatwiejszy w obsłudze niż interfejs NKJP i jego sformalizowany język zapytań.

Kategorie gramatyczne oraz części mowy w języku polskim i rosyjskim są podobne, ale nie identyczne. Tylko polszczyzna ma klasę bezosobników (form czasownikowych na *-no/-to*) i rodzaj męskoosobowy w liczbie mnogiej, tyko język rosyjski ma imiesłowy przymiotnikowe uprzednie czy krótkie formy przymiotników i imiesłowów biernych. W dodatku tager dla języka polskiego opiera się na fleksemach – klasach bardziej szczegółowych niż szkolne klasy części mowy. Wśród form tradycyjnego czasownika niedoświadczony użytkownik korpusu musi najpierw nauczyć się odpowiednich odrębnych tagów dla form czasu nieprzeszłego [fin], gerundiów [ger], predykatywów [pred] czy form typu *winien* [winien]. W wyszukiwaniu rosyjskim mamy zamiast klasy [fin] odpowiednio czasowniki w czasie teraźniejszym lub przyszłym, a pozostałych wymienionych klas nie ma w ogóle.

⁶ Wspomnijmy tu, nie nadając temu faktowi wielkiego znaczenia, że *Faraon* należał do ulubionych powieści Stalina. Wspomina o tym Józef Czapski (*Na nieludzkiej ziemi*), powołując się na rozmowy z I. Erenburgiem. Taką informację przynosi też hasło *Faraon* w rosyjskiej Wikipedii.

Trudność stanowi wyszukiwanie polskich form czasu przeszłego, które w języku rosyjskim jest tak samo łatwe, jak pozostałych czasów. Tager języka polskiego interpretuje formy czasu przeszłego 1. i 2. osoby zawsze jako ciąg imiesłowu przeszłego [praet] oraz aglutynacyjnej końcówki osobowo-liczbowej, czyli aglutynantu [aglt]: *-em, -eś, -śmy, -ście*. Jak wiadomo, te końcówki są w polszczyźnie ruchome, a rosyjski nie ma ich wcale. Dlatego, aby znaleźć polską formę 1. os. lp. w czasie przeszłym, trzeba w NKJP szukać ciągu [pos=praet & person=prim][pos=aglt & person=prim], a w rosyjskim korpusie narodowym wystarczy wybrać czas przeszły (nieodmienny przez osoby). W naszym korpusie przyjęliśmy drugi sposób wyszukiwania, ale program musi i tak znaleźć odpowiednie ciągi tagów, przy czym końcówki niestojące bezpośrednio po temacie czasu przeszłego nie zostaną znalezione.

Tagowanie morfosyntaktyczne NKJP skupia się na odmiennych częściach mowy. Wśród nieodmiennych rozróżnia się stosunkowo łatwe do wyodrębnienia według sąsiedztwa przyimki i spójniki oraz stojące przy czasownikach przysłówki. Tager rosyjski rozróżnia także partykuły.

Tager języka polskiego dezambiguuje tekst, tzn. na podstawie testów statystycznych oraz sąsiedztwa wyrazów wybiera dla danej formy homonimicznej najbardziej prawdopodobną interpretację gramatyczną, np. dla ciągu znaków *damy* tager wybierze interpretację D. lp., M./B./W. lm. od rzeczownika *dama* albo 1 os. lm. cz. nieprzeszłego czasownika *dać*. Rosyjski tager nie dezambiguje tekstów. W Rosyjskim Korpusie Narodowym dezambiguację przeprowadzono ręcznie dla niewielkiej części tekstów, w prezentowanym korpusie równoległym teksty nie były dezambiguowane. Dlatego zapytanie o polski leksem *dama* zwróci wśród wyników kontekst *odciski palców obu dam*, ale nie: *Dam panu znać*. Wyszukiwanie rosyjskiego leksemu *дaмa* da nie tylko *оbеux дам*, ale i nadmiarowe *Я вам дам знaть*.

6. Wyszukiwarki

Korpus został wyposażony w trzy wyszukiwarki:

1. WYSZUKIWANIE SŁÓW – umożliwia szukanie dokładnych form tekstowych bądź fragmentów form. Np. zapytanie o postaci **czyta** daje dostęp wyłącznie do tekstów zawierających po stronie polskiej formy 3. osoby czasu teraźniejszego czasownika *czytać*.

2. WYSZUKIWANIE ŁĄCZNE – umożliwia szukanie par tekstów, z których każdy spełnia określony warunek⁷. Gdy np. po stronie polskiej wpisujemy **czytani***, a po rosyjskiej – **читать**, otrzymamy dostęp do takich użyc polskiego gerundium *czytanie*, któremu w przekładach odpowiada (lub nie odpowiada, jeśli skorzystamy z okienka negacji) bezokolicznik *читать*. W pierwszym przypadku wśród 65 wyników otrzymamy m.in.:

To już lepiej ogłosić, że odbędzie się czytanie sztuki. [Komediantka, Reymont W.]	Тогда уж лучше объявить, что пьесу будут не играть, а читать . [Комедиантка, Реймонт В.]
Nie dokończywszy czytania Benedykt rzucił list na biurko. [Nad Niemnem, Orzeszkowa E.]	Бенедикт перестал читать и положил письмо на стол. [Над Неманом, Ожешко Э.]
Święte Oficjum nie zabrania czytania . Ani posiadania ksiąg. [Narrenturm, Sapkowski A.]	Святой Официум не запрещает читать книги и владеть ими. [Narrenturm, Сапковский А.]

Tabela 1. Wyniki wyszukiwania łącznego

W drugim przypadku (przy zanegowaniu rosyjskiego bezokolicznika) znajdziemy, np.:

W połowie czytania zmarszczył się, zmiął papier w rękę. [Król trędowaty, Kossak Zofia]	Однако же дочитав до конца, поморщился и скомкал листок в руке. [Король-крестоносец, Коссаk Зофья]
Nic mną tak dotychczas nie wstrząsnęło, jak czytanie przygód rozbitka na bezludnej wyspie. [Biały Jaguar, Fiedler Arkady]	Ничто ранее меня так глубоко не изумляло, как описания приключений Робинзона на необитаемом острове. [Белый ягуар, Фидлер Аркады]
Otóż wiadomo było, że pan dzięki temu, iż nie korzystał z umiejętności czytania ani pisania, miał fenomenalnie rozwiniętą pamięć wzrokową. [Cesarz, Kapuściński Ryszard]	Ведь известно, что император обладал феноменальной зрительной памятью. [Император, Капуцинский Рышард]

Tabela 2. Wyniki wyszukiwania łącznego z warunkiem

3. WYSZUKIWANIE MORFOLOGICZNE – pozwala dotrzeć do form zdefiniowanych pod względem gramatycznym. Można też szukać form dowolnych leksemów o zadanych parametrach, np. męskoosobowych mianownikowych form liczby mnogiej przymiotników w stopniu najwyższym. Znajdziemy wtedy m.in. *najstarsi*, *najgorsi*, *najsprytniejsi*, *najbliżsi*, *najlepsi*, *najulubiejsi* wraz z przekładami.

Wyszukiwanie możliwe jest w całym zbiorze tekstów lub w podzbiórach podzielonych na gatunki (fakt, fikcja, teksty prasowe, teksty prawne), przy

⁷ To szczególnie przydatny w korpusach równoległych rodzaj wyszukiwarki, której brakuje w wielu tego typu projektach, nawet zakrojonych na dużą skalę, jak np. w podkorpusie równoległym Narodowego Korpusu Języka Rosyjskiego.

czym w obrębie literatury pięknej w dwóch przedziałach czasowych: w dziełach napisanych przed rokiem 1945 lub po nim. Można też ograniczyć szukanie do języka oryginału.

We wszystkich wyszukiwarkach można stosować wyrażenia regularne, np. kropkę i gwiazdkę: *czyta*. zamiast *czytam, czytaj, czytać, czyta*.* zamiast *czytajcie, czytali*byście itd.

7. Przykłady zastosowań

7.1 Przykład pierwszy

Korpus równoległy może służyć jako pomocne narzędzie między innymi w przekładzie, przy budowaniu nowych słowników (jak się okazuje nie tylko dwujęzycznych, ale również jednojęzycznych typu onomazjologicznego) oraz przy weryfikacji danych słownikowych (i ocenie jakości przekładów). Pokażemy te możliwości na konkretnym przykładzie.

Zaimek *kilkanaście* nie ma w języku rosyjskim dokładnego ekwiwalentu. Trzy wielkie słowniki polsko-rosyjskie podsuwają odmienne rozwiązania przekładowe: *более десяти* (Hessen, Stypuła, 2004), *десятка полтора* (Wawrzyńczyk, red., 2005) i – znaczeniowo najmniej dokładne – *несколько* (Chwatow, Timoszuk, red., 2008). Sprawdźmy, jak ten problem przekładowy rozwiązują tłumacze-praktycy.

Wyszukiwarka morfologiczna wydaje 627 różnych kontekstów użycia w polskich tekstach leksemu *kilkanaście*. Rozkład ekwiwalentów proponowanych w wymienionych słownikach przedstawia się następująco: zaledwie w 8 przypadkach *kilkanaście* przetłumaczone zostało za pomocą wyrażenia *более десяти* (1,28 % wszystkich kontekstów), w 34 przypadkach tłumacze wybrali wyrażenie *десятка полтора* (6,22 % kontekstów) i aż w 186 przypadkach zdecydowali się na najmniej dokładny ekwiwalent – *несколько* (29,66 %).

W pozostałych 399 przypadkach tłumacze wykorzystali rozmaite inne wyrażenia z liczebnikami *десять, пятнадцать* i *двадцать* lub rzeczownikiem *десяток*, niekiedy z wykorzystaniem inwersji, będącej w ruszczyźnie sposobem wyrażania liczb przybliżonych. (por. zestawienie ekwiwalentów tekstowych, w którym w nawiasach w kolejności malejącej podana została liczba ich zastosowań, a wytłuszczonym drukiem zaznaczono ekwiwalenty słownikowe):

Ekwiwalenty przekładowe zaimka <i>kilkanaście</i> w Polsko-Rosyjskim Korpusie UW	Przykładów ogółem	%
более десяти , больше десяти	12	1,91
десять (z możliwą inwersją), десять с лишним, десять-пятнадцать, десять-пятнадцать (z możliwą inwersją), десять и больше, больше чем десять, едва ли десять, десять, а то и двадцать, десять-двенадцать	26	4,15
десяток, несколько десятков, десятка два, больше десятка, второй десяток, два десятка, с десяток, десятки, более десятка, десяток-другой, десяток с лишним, добрый десяток, около двух десятков, целые десятки, десяток с небольшим, десяток-два, какой-нибудь десяток, не меньше десятка, немногим больше десятка, около десятка, пара десятков, два-три десятка	133	21,22
десятка полтора (z możliwą inwersją)	48	7,66
пятнадцать (z możliwą inwersją), около пятнадцати, пятнадцать, пятнадцать-восемнадцать, пятнадцать-двадцать (z możliwą inwersją)	27	4,31
двадцать (z możliwą inwersją), каких-нибудь двадцать, около двадцати, почти двадцать, хоть двадцать, что-то около двадцати, не менее двадцати, двадцать-тридцать	41	6,54
Несколько	186	29,66
Inne	69	11,00
Brak przekładu	85	13,55
Razem	627	100%

Tabela 3. Ekwiwalenty przekładowe zaimka *kilkanaście* w Polsko-Rosyjskim Korpusie UW
Źródło: Opracowanie własne.

Jak wynika z zestawienia, w 85 przypadkach tłumacze podjęli decyzję o opuszczeniu ekwiwalentu (w tabeli rubryka: Brak przekładu), natomiast w 69 rozstrzygnęli problem w sposób indywidualny, mniej lub bardziej udany (rubryka: Inne). Dość skrajnym przypadkiem nieporadności wydaje się tłumaczenie wyrażenia *kilkanaście stóp wysokości* na *чуть ли не метр* (czyli dosłownie: 'bodaj metr')⁸.

Zebrane w korpusie wyrażenia mogą stać się podstawą opracowania słownika onomazjologicznego, w tym przypadku obejmującego środki wyrażania wybranych liczb przybliżonych.

Na szczególną uwagę i uwzględnienie w przyszłych słownikach polsko-rosyjskich zasługują niektóre odpowiedniości przekładowe pozostające dotąd poza polem widzenia leksykografów. Chodzi mianowicie o połączenia rozważanego zaimka z liczebnikiem głównym, typu *sto kilkaścicie*, *dwieście kilkaścicie*

8 Por.: Z гашzczу кактусов, доходящих некийдо до *kilkunastу стоп wysokości*, долатыwały мне весоле щчебиоты и погвизды птаков. [A. Fiedler: *Бялы Jaguar*] – Из зарослей кактусов, достигавших порой чуть ли не метра в высоту, доносился веселый щебет птиц. [Белый ягуар].

itd., których przekład wymaga sięgnięcia po środki, jakich żaden z istniejących słowników nie proponuje. Spójrzmy na przykłady:

<u>Stu kilkunastu</u> nas dołożyło ramienia. [Popioły, Żeromski S.]	Сто с лишним человек помучились с ним. [Пепел, Жеромский С.]
W parę tygodni czasu z naszego batalionu, który tysiąc miał ludzi, zostało nas <u>stu kilkunastu</u> . [Popioły, Żeromski S.]	Через две недели из батальона, в котором было тысяча человек, осталось нас сто с небольшим . [Пепел, Жеромский С.]
Zaledwie trzeciego dnia zebrała się koło niego wataha złożona z <u>dwustu kilkunastu</u> mołojców (...). [Ogniem i mieczem, Sienkiewicz H.]	Только на третий день около него собралась ватага казаков из двухсот с небольшим человек (...). [Огнем и мечом, Сенкевич Г.]

Tabela 4. Ekwiwalenty tłumaczeniowe zaimka *kilkanaście* w Polsko-Rosyjskim Korpusie UW

Liczne tego rodzaju odpowiedniości upoważniają do zaproponowania następującej reguły przekładowej, którą można włączyć do słownika dwujęzycznego:

sto (dwieście, trzysta *itd.*) **kilkanaście** сто (двести, триста *и т.д.*)
с небольшим, сто (двести, триста *и т.д.*) с лишним.

Jak pokazują materiały korpusu, zasługującym na uwagę chwytem tłumaczeniowym, dość często stosowanym przy przekładzie połączeń zaimka *kilkanaście* z rzeczownikami oznaczającymi jednostki czasu (*sekunda, minuta, godzina, dzień, miesiąc*), jest oddawanie ich w ruszczyźnie za pomocą konstrukcji z rzeczownikami oznaczającymi odpowiednio większe odcinki czasu (*полминуты, четверть часа, сутки, неделя, год*), por. np.:

– Teraz Rührtgard był wyraźnie zablokowany i minęło <u>kilkanaście sekund</u> , zanim wyrzucił z siebie: – Jedynym zresztą, jakiego miałem i mam... [Widma w mieście, Krajewski M.]	– У Рютгарда вдруг будто язык отнялся, прошло чуть ли не полминуты , прежде чем ему удалось договорить. – Единственный друг за всю мою жизнь... [Призраки Бреслау, Краевский М.]
Po <u>kilkunastu minutach</u> przechadzki Nikodem wybrał jedną z nich, znajdującą się w gęstym cieniu, i usiadł. [Kariera Nikodema Dyźmy, Dołęga-Mostowicz T.]	Побродив с четверть часа , Никодим выбрал одну из них, в густой тени, и сел. [Карьера Никодима Дызьмы, Доленга-Мостович Т.]
Od <u>kilkunastu godzin</u> straciliśmy panowanie nad sterem. [Biały Jaguar, Fiedler A.]	Почти сутки мы без руля. [Белый ягуар, Фидлер А.]
W Minturnae, gdzie znów śpiewał publicznie, zabawił <u>dni kilkaście</u> , a nawet znów się namyślał, czy nie wrócić do Neapolu (...) [Quo Vadis. Powieść z czasów Nerona, Sienkiewicz H.]	В Минтурнах, где он опять пел при публике, он провел недели две и даже стал подумывать о том, не вернуться ли в Неаполис (...) [Камо грядеши, Сенкевич Г.]
<u>Kilkanaście miesięcy</u> przynajmniej mieć musiało, skoro już mówiło i przedmioty rozpoznawać mogło (...). [Cham, Orzeszkowa E.]	Ребенку было уже, вероятно, больше года , так как он уже говорил и умел различать предметы (...). [Хам, Ожешко Э.]

Tabela 5. Ekwiwalenty tłumaczeniowe zaimka *kilkanaście* w Polsko-Rosyjskim Korpusie UW

I w tym przypadku można się pokusić o zaproponowanie ogólniejszej reguły przekładowej.

7.2 Przykład drugi

Wyszukajmy te zdania rosyjskie, w których występuje trudny do zrozumienia dla Polaków przymiotnik *российский* (zapisany za pomocą wyrażen regularnych *российск.*|Российск.**), ale takie, których odpowiedniki polskie nie zawierają przymiotnika *rosyjski* (*rosyjs.*|Rosyjs.**). Są 32 takie zdania, czyli 15% ze 195 wszystkich wystąpień przymiotnika *российский*. Analogiczne porównanie dla przymiotnika *русский* bez odpowiednika *rosyjski* po stronie polskiej daje 650 wyników, czyli 37% z 1757 wszystkich wystąpień przymiotnika *русский*. Przymiotnik *российский* jest częściej niż *русский* zastępowany w polskim przekładzie przez przydawkę dopełniaczową *Rosji* (np. *российское влияние – wpływy Rosji, российский премьер – premier Rosji*), bywa też pozostawiany bez odpowiednika (np. „Польскую” часть захватили немцы, а после войны она вместе с другими трофеями была перевезена в Москву и теперь находится в Российском государственном военном архиве. [Коммерсантъ, 2011/35] – „Polską” część w czasie okupacji zagarnęli Niemcy, po II wojnie światowej została przewieziona do Moskwy i zdeponowana w Państwowym Archiwum Wojskowym. [Forum 2011/40]), ale w 85% procentach jest tłumaczony na neutralny przymiotnik *rosyjski* (np. *Теперь важно дождаться реакции российских официальных лиц (...)* [politcom.ru 29.07.2011] – *Teraz czekamy na reakcję rosyjskich czynników oficjalnych (...)* [Forum 2011/31]). Przymiotnik *русский* jest częściej tłumaczony inaczej niż *rosyjski* (w co trzecim użyciu), ponieważ w opisach rzeczywistości dawnej i w tekstach stylizowanych potocznie odpowiada mu polski przymiotnik *ruski*, ponadto w użyciu rzeczownikowym *русский* to po polsku *Rosjanin*.

8. Znaczenie korpusu i perspektywy jego rozwoju

Korpus Polsko-Rosyjski UW jest efektem zamkniętego projektu badawczego, ale nie wyklucza to jego rozwoju i rozbudowy. Już w tej chwili spełnia on ważną funkcję w dydaktyce rusycystycznej w Polsce (np. Kuratczyk, 2013, 2014; Kuratczyk, 2015) oraz w badaniach naukowych (np. Łaziński, 2015). Ze względu na cechy szczególne korpusów równoległych, tj.: bogaty opis bibliograficzny tekstów literackich czy też udział tekstów prasowych o tematyce lokalnej, korpus mógłby stać się cennym składnikiem większego polskiego projektu korpusów równoległych lub nawet podstawą takiego projektu. Przyszłość korpusu zależy w pewnym stopniu od przyszłości zainteresowania językiem i kulturą rosyjską w Polsce. Jego twórcy mają nadzieję, że przyczynili się do zwiększenia tego zainteresowania.

Literatura

- CHWATOW, Sergiusz, TIMOSZUK, Mikołaj (red.) (2008): *Wielki słownik rosyjsko-polski, polsko-rosyjski*. Warszawa: REA.
- HESSEN, Dymitr, STYPUŁA, Ryszard (2004): *Wielki słownik polsko-rosyjski, tom 1-2*. Warszawa: Wiedza Powszechna.
- KURATCZYK, Magdalena (2013): Polsko-rosyjskie korpusy równoległe i przykład ich wykorzystania w gramatyce konfrontatywnej. *Komunikacja Specjalistyczna. Tropy lingwistyki* 6, 44–62.
- KURATCZYK, Magdalena (2014): Параллельные польско-русские корпуса в обучении переводу юридического текста. *Komunikacja Specjalistyczna. Languages for Special Purposes – a Tool of Modern Science* 7, 111–128.
- KURATCZYK, Magdalena (2015): Параллельные корпуса с русским и польским языками и их использование в сопоставительной лингвистике и лингводидактике. *Русский язык и литература в пространстве мировой культуры, Материалы XIII конгресса МАПРЯЛ в 15 томах, Гранада, Испания, 13-20 сентября 2015*, Granada: Universidad de Granada, XI, 153–157.
- ŁAZIŃSKI, Marek, KURATCZYK, Magdalena (2015): Польско-русский параллельный корпус Варшавского университета и его использование в лингвистическом исследовании. W: Dorota SZUMSKA, Krzysztof OZGA, (red.). *Язык и метод. Русский язык на грани методологического срыва*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 95–103.
- ŁAZIŃSKI, Marek, KURATCZYK, Magdalena, OREKHOV, Boris, SŁOBODJAN, Elena (2012): The Polish-Russian Parallel Corpus and Its Application in the Linguistic Analysis. *Prace Filologiczne* LXIII, 209–218.
- ŁAZIŃSKI, Marek (2015): Функции инфинитива в польском и русском языках на фоне корпусов. W: Dorota SZUMSKA, Krzysztof OZGA, (red.). *Язык и метод. Русский язык на грани методологического срыва*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 157–168.
- WAWRZYŃCZYK, Jan (red.) (2005): *Wielki słownik polsko-rosyjski*. Warszawa: PWN.

