



ANDRZEJ CHARCIAREK

 ORCID: <https://orcid.org/0000-0001-5651-5730>

Uniwersytet Śląski w Katowicach

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

PARALLEL CORPUS AND EQUIVALENCING

The article is devoted to possibilities of using the InterCorp Polish-Russian parallel corpus in the process of searching for translations equivalents. It also presents resources and functionalities of the InterCorp, and concentrates on how it can be used by Russian philologists. It is increasingly useful in linguistic and translational research. A particular focus is put on the process of equivalencing i.e. searching for and establishing equivalents and a significant role of the parallel corpus which allows for finding appropriate translation solutions that sometimes can pose a problem for a single translator or lexicographer. The article also presents a deeper reflection on the significance and prospects of empirical research in the contemporary linguistics.

Badań na gruncie językoznawstwa korpusowego prowadzonych w polskim środowisku rusycystycznym nie znajdziemy wiele. Najlepszym tego dowodem jest monografia *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*¹, stanowiąca w zamierzeniu jej redaktora, Wojciecha Chlebdy, „obudowę przysłownikową *Podręcznego idiomatykonu polsko-rosyjskiego*”², a poświęcona zastosowaniu korpusów w badaniach lingwistycznych. Wśród autorów poszczególnych rozdziałów znaleźli się zarówno twórcy korpusów, jak i ich użytkownicy: Milena Hebal-Jezierska, Piotr Fliciński, Piotr Pęczik, Tadeusz Piotrowski, Łukasz Grabowski, Mirosław Bańko, Marek Iwanowski, Piotr Wierchoń i Jan Wawrzyńczyk. Już sama analiza składu autorów dowodzi, że rusycyści są

¹ W. Chlebda (red.), *Na tropach tłumaczeń. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opoleńskiego, Opole 2013.

² W. Chlebda, *Korpusologia użytkowa dla początkujących i zaawansowanych*, w: W. Chlebda (red.), *Na tropach tłumaczeń...*, s. 7–15.

reprezentowani jedynie przez Jana Wawrzyńczyka (redaktora monografii, rusycystę, pomijam). I choć od wydania monografii minęło wiele lat, to niewiele się zmieniło. W dalszym ciągu badania korpusowe szerzej i intensywniej rozwijane są w polskim środowisku anglistycznym, polonistycznym czy bohemistycznym. Na taki stan rzeczy składa się wiele czynników, jak choćby tworzenie korpusów narodowych i zaangażowanie w nie przedstawiciele poszczególnych filologii³. Nie sposób również nie dostrzegać znaczenia tradycji badań korpusowych, które powstały i rozwijały się głównie w środowisku anglistycznym⁴.

W niniejszym tekście, choć będzie on głównie poświęcony roli korpusu równoległego InterCorp⁵ w procesie ekwiwalentyzacji⁶, przywołam na wstępie jeszcze dwa korpusy, które — jak zakładałam — są najbardziej znane w środowisku rusycystycznym.

Pierwszy z nich to Polsko-Rosyjski i Rosyjsko-Polski Korpus Równoległy Uniwersytetu Warszawskiego⁷ zawierający 30 milionów segmentów. Należy dodać, że za segmenty uważa się nie tylko tradycyjnie rozumiane formy wyrazowe, lecz także aglutynacyjne morfemy osobowo-liczbowe czasownika. Na jego zawartość składają się teksty literackie, prasowe, religijne i prawne. Jego zaletą z pewnością są przekłady bezpośrednie polsko-rosyjskie i rosyjsko-polskie dodatkowo uzupełnione polskimi i rosyjskimi przekładami tekstów literackich w języku angielskim, niemieckim i francuskim. Proporcje ułożyły się następująco: 50% oryginałów polskich, 33% oryginałów rosyjskich i 17% oryginałów z języków trzecich. Można różnie oceniać proporcje poszczególnych tekstów, zwłaszcza 90-procentowy udział tekstów literackich może wydawać się dla jednych użytkowników zaletą, dla drugich — wadą. Ważną informacją jest ograniczony dostęp do 34% tekstów, które są dostępne tylko dla współpracowników korpusu. Z pozostałych 66% tekstów można korzystać bez logowania.

³ Należy tu wspomnieć także o projekcie *Wielkiego słownika języka polskiego PAN* powstającym głównie w oparciu o teksty Narodowego Korpusu Języka Polskiego.

⁴ Nie przypadkiem pierwszy polski podręcznik językoznawstwa korpusowego powstał w środowisku anglistycznym: B. Lewandowska-Tomaszyk (red.), *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.

⁵ <https://intercorp.korpus.cz>.

⁶ To właśnie ten termin znalazł się w tytule, ponieważ — moim zdaniem — bardzo dobrze oddaje specyfikę pracy z korpusem równoległym.

⁷ <http://pol-ros.polon.uw.edu.pl>.

Drugi, polsko-rosyjski korpus równoległy, stanowiący integralną część Narodowego korpusu języka rosyjskiego⁸, liczy ponad 6 milionów słów. Wśród tekstów znajdziemy sporo polskiej i rosyjskiej literatury klasycznej (Sienkiewicz, Prus, Briusow, Sołogub), ale także nielicznych autorów współczesnych (Sapkowski, Chmielewska). Zawartość polsko-rosyjskiego korpusu równoległego uzupełniają teksty prawne i artykuły prasowe. Niestety niewielkie zasoby tego korpusu praktycznie uniemożliwiają pozyskiwanie odpowiedniej liczby poświadczeń do analizy korpusowej.

Trzeci, to wspomniany polsko-rosyjski korpus równoległy InterCorp stanowiący integralną część Narodowego korpusu języka czeskiego (Český národní korpus), regularnie wzbogacany o nowe teksty. InterCorp jest korpusem wielojęzycznym (ostatnia wersja 13. gromadzi 40 języków), w którym język czeski zajmuje pozycję uprzywilejowaną. Nie sposób nie wspomnieć o mocnej pozycji języka angielskiego, który poza czeskim jest drugim językiem interfejsu. Każdy czeski tekst w korpusie posiada odpowiednik, którym może być obcojęzyczny oryginał lub jego tłumaczenie na język obcy. Najpełniejszy opis InterCorp znajdziemy w *Praktycznym przewodniku po korpusie równoległym InterCorp*⁹ Elżbiety Kaczmarek i Alexandra Rosena.

Każdy rok przynosi nową wersję korpusów poszczególnych języków, których zasoby są systematycznie powiększane. W momencie powstawania niniejszego tekstu użytkownicy mogą korzystać z 13. wersji udostępnionej w 2020 roku. W wersji tej wielkość korpusu rosyjskiego po raz pierwszy przekroczyła liczbę 20 milionów słów, a polskiego — zbliżyła się do 90 milionów słów. Dowodem na sukcesywnie rosnące rozmiary korpusu polskiego i rosyjskiego jest porównanie ich ostatnich trzech wersji (wielkości podane w liczbach bezwzględnych):

— polski moduł InterCorp: 11 wersja — 86 230 000, 12 wersja — 87 567 000, 13 wersja — 88 161 000;

— rosyjski moduł InterCorp: 11 wersja — 18 282 000, 12 wersja — 19 560 000, 13 wersja — 20 078 000.

Analiza zasobów pokazuje, że oba korpusy — polski i rosyjski — powiększają swoje rozmiary dzięki włączeniu do nich tekstów beletrystycznych. Wielkość poszczególnych kolekcji z przyczyn często

⁸ <https://ruscorpora.ru>.

⁹ E. Kaczmarek, A. Rosen, *Praktyczny przewodnik po korpusie równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowiańskich*. M. Hebal-Jeziarska (red), Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa, s. 207–231.

obiektywnych pozostaje praktycznie niezmienna. Zjawisko to dotyczy kolekcji, które nie mogą być powiększane, ponieważ tworzą zamknięte zbiory tekstów. Przykładem tego jest kolekcja tekstów publicystycznych z nieistniejącego już serwisu www.presseurop.eu. Analogiczna sytuacja ma miejsce w przypadku zbioru zawierającego zapisy obrad Parlamentu Europejskiego z lat 2007–2011 z korpusu Europarl.

Co zrozumiałe, korzystanie z poszczególnych kolekcji zależy od tego, czy są obecne w obu zestawianych ze sobą językach. To one tworzą korpus równoległy tekstów sobie odpowiadających, mogący skutecznie posłużyć do pracy translatorskiej czy leksykograficznej.

O tym, jak istotne znaczenie ma wielkość i powiększanie zasobów korpusowych i, co się z tym bezpośrednio wiąże, przydatność korpusu do wspomnianych badań, świadczą liczby poświadczeń tych samych wyrażeń¹⁰ w poszczególnych jego wersjach, na przykład:

wyrażenie	InterCorp	wersja 13	Wersja 12	Wersja 11
<i>złożyć podanie</i>	pol.-ros. pol.-czes.	12 55	11 54	— ¹¹ 2
<i>spokojna głowa</i>	pol.-ros. pol.-czes.	12 81	12 81	12 79
<i>niedźwiedzia przysługa</i>	pol.-ros. pol.-czes.	1 20	1 19	— 18 ¹²
<i>dyrektor finansowy</i>	pol.-ros. pol.-czes.	50 59	50 59	50 59
<i>rodzona matka</i>	pol.-ros. pol.-czes.	2 17	2 16	5 29
<i>walić głową w mur</i>	pol.-ros. pol.-czes.	4 12	4 12	3 12

Jak widać, choć każda nowa wersja InterCorp jest większa od poprzedniej, to wyszukiwanie określonych wyrażeń w najnowszej wersji nie oznacza automatycznie większej liczby poświadczeń. Dobrą ilustracją tego zjawiska jest wyrażenie *rodzona matka*, które w polsko-czeskiej 11. wersji posiada znacznie więcej wyników (29), niż w naj-

¹⁰ Choć są wyjątki.

¹¹ Wyrażenie *złożyć podanie* nie ma poświadczeń w wyszukiwaniu z atrybutem *lemat*, natomiast za pomocą atrybutu *forma wyrazowa (word)* uzyskujemy 4 poświadczenia.

¹² W wersji 11. InterCorp wyszukiwanie z wykorzystaniem atrybutu *lemat* ciągu *niedźwiedzi przysługa* nie przynosi żadnego poświadczenia. Uzyskane 18 wyników to rezultat wyszukiwania *niedźwiedź przysługa*. Najpewniej przyczyną takiego stanu rzeczy był błąd w anotacji, który już w kolejnej 12. wersji został usunięty.

nowszej 13. (17). Fakt ten może skłaniać użytkownika do korzystania ze starszych wersji InterCorp, tym bardziej że taka możliwość istnieje, ponieważ wszystkie wersje, począwszy od 6., są wciąż dostępne. Znaczenie wielkości korpusu dobrze obrazuje liczba poświadczeń lematu *książka* czy *wziąć*, jednostek ogólnoodmianowych, w poszczególnych wersjach korpusu:

wyraz	InterCorp	wersja 13	wersja 12	wersja 11
<i>książka</i>	pol.-ros.	3380	3129	2788
	pol.-czes.	16299	16201	15836
<i>wziąć</i>	pol.-ros.	7111	6882	6513
	pol.-czes.	41404	41128	39554

Uzyskany materiał empiryczny pozwala na sformułowanie kilku uwag natury ogólniejszej.

Po pierwsze, im większy korpus, tym więcej poświadczeń. Zjawisko to widać zarówno między zestawianymi językami (polsko-rosyjski, polsko-czeski), jak i ich kolejnymi wersjami polsko-rosyjskiego czy polsko-czeskiego InterCorp.

Po drugie, skala wzrostu korpusu polsko-rosyjskiego nie dotyczy — co znamienne — w równym stopniu wszystkich notowanych w nim jednostek. Można wykazać, że dodanie napisów filmowych spowoduje wzrost liczby rejestrowanych w nim jednostek potocznych, dodanie tekstów administracyjnych — typowych dla nich sformułowań.

Po trzecie, korpus InterCorp nie zwiększa swoich rozmiarów w sposób proporcjonalny; między poszczególnymi wersjami mogą być mniejsze lub większe różnice. Przeprowadzona sonda odzwierciedla w zasadzie proporcje między przywołanymi wyżej wielkościami omawianych korpusów — większy wzrost można zaobserwować między wersją 11. a 12. niż między 12. a 13.

Niekorzystnym zjawiskiem jest to, że oprócz wiązanych ręcznie tekstów beletrystycznych, tworzących rdzeń (czes. jądro), zasobami, mogącymi służyć do ustanowienia ekwiwalencji polsko-rosyjskiej pozostają tylko dwie kolekcje wiązane automatycznie: napisy filmowe z amatorskiego serwisu OpenSubtitles oraz przekłady Biblii. Trzecia kolekcja tekstów rosyjskich dostępna w InterCorp — teksty prasowe z platformy Project Syndicate, nie może być wykorzystana, ponieważ brakuje jej polskich odpowiedników. Korpus polsko-rosyjski dysponuje więc tekstami ograniczonymi pod kątem ich zróżnicowania gatunkowego; wyraźnie brakuje tekstów prawnych, publicystycz-

nych czy ekonomicznych. O tym, jaki ma to wpływ na wyszukiwanie w korpusie, można się przekonać, porównując polsko-czeski InterCorp z polsko-rosyjskim. Kwerenda najbardziej typowych terminów prawnych i ekonomicznych przynosi następujące rezultaty (liczby poświadczeń):

wyrażenie	Polsko-rosyjski InterCorp	Polsko-czeski InterCorp
<i>przedmiot umowy</i>	0	145
<i>umowa o pracę</i>	2	439
<i>zabrania się</i>	3	211
<i>decyzja w sprawie</i>	0	2582
<i>wzrost gospodarczy</i>	0	2621
<i>obligacje państwowe</i>	2	99
<i>lokata terminowa</i>	0	6
<i>niniejszym informować</i>	0	13
<i>zapytanie ofertowe</i>	1	2
<i>rynek międzybankowy</i>	0	21

Powyższe wartości dobrze dokumentują wagę obecności tekstów prawnych i ekonomicznych w polsko-czeskim InterCorp, dzięki którym pozyskiwane są większe lub mniejsze liczby poświadczeń. Jednocześnie brak tego typu tekstów w korpusie polsko-rosyjskim uniemożliwia praktycznie pozyskanie jakichkolwiek wyników. W praktyce oznacza to, że użytkownik nie jest nawet naprowadzany na ewentualny ekwiwalent czy ekwiwalenty. Źródło pochodzenia poszczególnych poświadczeń można określić za pomocą opcji *typy tekstów* (czes. *typy textů*):

wyrażenie	Korpus	Liczba poświadczeń	Core	Acquis	Euro-parl	Press-Europ	Sub-titles
<i>przedmiot umowy</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	145	10	127	8	–	–
<i>umowa o pracę</i>	pol.ros.	2	2	–	–	–	–
	pol.-czes.	439	15	298	114	10	3
<i>zabrania się</i>	pol.-ros.	3	1	–	–	–	2
	pol.-czes.	211	10	154	16	7	24
<i>decyzja w sprawie</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	2582	18	1811	693	50	10
<i>wzrost gospodarczy</i>	pol.ros.	–	–	–	–	–	–
	pol.-czes.	2621	1	540	1794	280	6

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

<i>obligacje państwowe</i>	pol.-ros.	2	2	–	–	–	–
	pol.-czes.	99	3	39	12	43	2
<i>lokata terminowa</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	6	–	5	–	1	–
<i>niniejszym informować</i>	pol.-ros.	1	1	–	–	–	–
	pol.-czes.	13	2	8	2	–	1
<i>zapytanie ofertowe</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	2	1	1	–	–	–
<i>rynek międzybankowy</i>	pol.-ros.	–	–	–	–	–	–
	pol.-czes.	21	–	8	10	3	–

Jak ważna jest weryfikacja pozyskanego materiału przekonuje rejestracja przez wyszukiwarkę Kontext poświadczeń, mających odmienną semantykę. Wprowadzenie zapytania *zabraniać się* z atrybutem *lemat* pozwala na uzyskanie wyników w znakomitej większości tego zwrotu, który był przewidywany, czyli *zabrania się* (od czasownika *zabraniać się*). Jednakowoż zwrot ten, typowy dla tekstów administracyjnych typu rozporządzenie czy zarządzenie, przybiera taką samą postać, jak *zabrania się*, będący formą rzeczownika odsłownego *zabranie się* w dopełniaczu, np. *dasz mi szansę zabrania się na samolot*. Przykład ten (z części polsko-czeskiej InterCorp) dowodzi obowiązku sprawdzenia pozyskanych poświadczeń, ponieważ domysł czy intuicja użytkownika mogą okazać się zawodne.

Na korpus polsko-rosyjski, zresztą na pozostałe w InterCorp również, składają się głównie tłumaczenia, czyli przekład bezpośredni polsko-rosyjski lub rosyjsko-polski ma miejsce rzadko.

Właśnie InterCorp ze względu na swoją unikatowość skłania do głębszej refleksji nad miejscem korpusu równoległego w badaniach lingwistycznych, a szczególnie w procesie ustalania ekwiwalentów przekładowych. I to z dwu przyczyn.

Pierwsza jest oczywista. Wykorzystanie korpusów równoległych, zresztą nie tylko tego rodzaju, w rusycystycznych badaniach językoznawczych jest niewielkie. Dobrze więc byłoby ten stan rzeczy zmienić.

Druga przyczyna, niejako wynikająca z pierwszej, to wskazanie jego zalet odnośnie do pracy translatorskiej i leksykograficznej, nie mówiąc o jego przydatności w badaniach glottodydaktycznych czy socjolingwistycznych. O każdym z wymienionych pól zastosowań można napisać odrębnie i obszernie zarazem. Tu skupimy się wyłącznie na procesie poszukiwania ekwiwalentów przekładowych, czyli na procesie, który Wojciech Chlebda nazywa ekwiwalentyzacją. Jego

zdaniem, przyjęcie perspektywy patrzenia na pracę tłumacza czy leksykografa przekładowego wymusza dodanie do pojęć *ekwiwalent* (na gruncie leksykografii — jednostki języka) i *ekwiwalencja* (relacja wzajemnej odpowiedniości między jednostkami) pojęcia trzeciego, wspomnianego już pojęcia ekwiwalentyzacji, rozumianej jako proces poszukiwania i dochodzenia do jednostek pozostających w danej relacji¹³. Teksty korpusowe są chyba jednym z najlepszych narzędzi, aby ten proces obserwować. To właśnie korpusy równoległe najlepiej obrazują to, co Chlebda nazywa „stopniowym osaczeniem odpowiednika przekładowego”¹⁴. W rzeczy samej korpus równoległy może stanowić szczególnego rodzaju „słownik propozycji translatorskich”¹⁵ ze wszystkimi słabościami i ograniczeniami, ale też niewątpliwymi zaletami. Wybory translatorskie znajdujące się w profesjonalnych pracach leksykograficznych są najczęściej autorstwa jednego czy kilku-kilkunastu specjalistów leksykografów. Propozycje w korpusie równoległym natomiast to rozwiązania heterogeniczne — przedstawiciele różnych generacji, grup społecznych i zawodowych, w różnym wieku i różnej płci. W odróżnieniu od indywidualnego idiolektu konkretnego słownikarza czy tłumacza, w korpusie równoległym znajdujemy setki czy tysiące poświadczeń, rejestrujących idiolekty ich autorów. Wszystko to stwarza zupełnie nową sytuację, w której możliwe jest wskazanie takich odpowiedniości przekładowych, które przekraczają możliwości jednego użytkownika języka, nawet biorąc pod uwagę jego ponadstandardowe zdolności intelektualne i językowe. To ograniczenie dotyczy bez wyjątku każdego użytkownika, który korzysta „z takich możliwości, jakie oferuje mu jego język, ale w ramach tych możliwości sam wybiera sposób obrazowania”¹⁶.

Wielkość i zróżnicowanie tekstów korpusu są niezbędne dla skutecznej analizy statystycznej i, co z tym się bezpośrednio wiąże, ekwiwalentyzacji. W przypadku polsko-rosyjskiego InterCorp — jak już stwierdziłem i zilustrowałem na konkretnych przykładach — o dużych możliwościach nie może być mowy, ponieważ i jedna, i druga wartość nie jest duża. Rzecz jasna, to wcale nie przesądza o jego nieprzydatności w procesie tłumaczenia określonych typów tekstów.

¹³ W. Chlebda, *Ekwiwalencja i ekwiwalenty: między słownikiem a tekstami*, w: W. Chlebda (red.), *Na tropach translatów...*, s. 41.

¹⁴ Tamże, s. 42.

¹⁵ Tamże.

¹⁶ E. Tabakowska, *Gramatyka i obrazowanie. Wprowadzenie do językoznawstwa kognitywnego*, Wydawnictwo Oddziału PAN, Kraków 1995, s. 57.

Wszystko zależy od tego, jakie jednostki mają być tłumaczone. Zasoby korpusowe, czyli gatunki tekstów zgromadzonych w korpusie, decydują o jego przydatności w wyszukiwaniu ekwiwalentów określonych translandów. Nieobecność w korpusie tekstów prawnych czy ekonomicznych automatycznie spowoduje, że użytkownik nie znajdzie w nim np. skonwencjonalizowanych sformułowań typowych dla decyzji administracyjnej czy rozporządzenia. Brak tego typu gatunków administracyjnych cechujących się własnymi sposobami realizacji tekstowych oraz wyspecjalizowanymi formami językowymi skąpe praktycznie każdą kwerendę korpusową na niepowodzenie. Wydaje się, że wiedza ta, choć oczywista, nie jest powszechna. Niewielka jest świadomość faktu, że w przypadku korpusu równoległego stworzenie odpowiednio dużego korpusu jest prawdziwym wyzwaniem; pozyskiwanie tekstów napotyka różnego rodzaju trudności, choćby wynikające z ochrony praw autorskich. Porównanie zasobów korpusu polsko-rosyjskiego i polsko-czeskiego InterCorp dobrze zjawisko to ilustruje. Jeśli pierwszy zawiera teksty beletrystyczne i dwie kolekcje (Biblia, napisy filmowe z platformy OpenSubtitles), to drugi dysponuje, oprócz literatury pięknej, pięcioma kolekcjami: aktami prawnymi Unii Europejskiej z korpusu *Acquis Communautaire*, zapisami posiedzeń Parlamentu Europejskiego z lat 2007–2011, tekstami publicystycznymi z serwisu www.presseurop.eu, napisami filmowymi z platformy *Opensubtitles* i Biblią.

Aby uzmysłwić sobie znaczenie wielkości korpusu i różnorodności tekstów w nim zgromadzonych, można porównać liczbę poświadczeń wybranych jednostek wielowyrazowych w układzie polski – rosyjski i polski – czeski¹⁷:

wyrażenie	InterCorp polsko-rosyjski	InterCorp polsko-czeski
<i>złożyć podanie</i>	12 poświadczeń	55 poświadczeń
<i>spokojna głowa</i>	12 poświadczeń	88 poświadczeń
<i>niedźwiedzia przysługa</i>	1 poświadczenie	20 poświadczeń
<i>dyrektor finansowy</i>	50 poświadczeń	59 poświadczeń
<i>rodzona matka</i>	2 poświadczenia	17 poświadczeń
<i>walić głową w mur</i>	4 poświadczenia	12 poświadczeń

Już wstępna analiza danych korpusowych pokazuje, że w wielu przypadkach przeprowadzenie analizy korpusowej i wskazanie par

¹⁷ Wyniki dla wyszukiwania z atrybutem *lemat*.

ekwiwalentów nie jest możliwe; poświadczeń jest zbyt mało. Widać dużą różnicę w liczbie pozyskanych poświadczeń między korpusem równoległym polsko-czeskim a polsko-rosyjskim na korzyść tego pierwszego. Jednocześnie zestawienie poświadczeń dokumentuje nie tylko znaczenie wielkości korpusu, ale także zróżnicowanie gatunkowe zgromadzonych w nim tekstów.

Dla pozyskania jak największej liczby poświadczeń wykorzystałem atrybut *lemat (lemma)*, który pozwala uwzględnić wszystkie formy utworzone od wpisanych leksemów *walić głowa w mur*, np. *walisz głową w mur*, *walą głową w mur*¹⁸.

O tym, jak istotne jest zbudowanie zapytania i weryfikacja automatycznie pozyskanych danych korpusowych, przekonuje poszukiwanie rosyjskich i czeskich ekwiwalentów polskiego wyrażenia *spokojna głowa*. W *Wielkim słowniku języka polskiego* IJP PAN jest ono definiowane następująco: ‘*pot.* wypowiedane w celu uspokojenia osoby, która się czymś martwi’¹⁹. Pozostawienie atrybutu *lemat* i wpisanie w wyszukiwarkę wyrazów hasłowych *spokojny głowa* umożliwi pozyskanie 12 par polsko-rosyjskich i 88 polsko-czeskich. Niemniej po bardziej szczegółowym przyjrzeniu się pozyskanym poświadczeniom okazuje się, że wśród zgromadzonego materiału znajdują się takie, o które w analizie nie chodziło:

Potem już ze spokojną głową będę mógł znaleźć jakieś odpowiednie zajęcie, które by mi dało przyzwoity zarobek i pozostawiło dosyć wolnego czasu na pisanie;
Wyjedziesz ze spokojną głową;
A ja upodobałem sobie ciebie i tę młodą kobietę, toteż powiedziałem tylko słówko i możecie już mieć spokojną głowę.

Jak widać, przytoczone poświadczenia (*ze spokojną głową*, *mieć spokojną głowę*), rejestrują inne znaczenie, nienotowane w słownikach, a które można ująć jako ‘nie zaprzętać sobie głowy innymi problemami’.

W przypadku badania wyrażenia *spokojna głowa*, występującego wyłącznie w tej jednej utrwalonej postaci, lepszym rozwiązaniem jest zatem użycie atrybutu *forma wyrazowa (word)*, który spowoduje wyszukanie jedynie wyrażenia *spokojna głowa* (81 poświadczeń).

Inną strategię wyszukiwawczą należałoby zastosować wobec wyrażenia *niedźwiedzia przysługa*, które może być w tekście poprzedzone

¹⁸ Kontext wyszukuje także przykłady z rzeczownikiem odczasownikowym, tj. *walenie głową w mur*.

¹⁹ <https://wsjp.pl/haslo/podglad/18949/spokojna-glowa>.

czasownikami *wyświadczyć*, *wyrządzić*, *oddać* czy *zrobić*²⁰. Użycie atrybutu *lemat* zapewnia maksymalne wykorzystanie zasobów korpusowych, choć w przypadku korpusu rosyjskiego niewiele to zmienia (1 poświadczenie). Jeszcze raz wielkość tego korpusu uniemożliwia jakąkolwiek analizę korpusową.

Dodatkowo pośród pozyskanych poświadczeń mogą się znajdować (i się znajdują) pary przekładowe, które takimi nie są. Wskazywane przez Kontext, podstawową wyszukiwarkę InterCorp, translaty nie mają niekiedy żadnego związku z translandem. Przyczyną takiego stanu rzeczy najczęściej jest automatyczne wiązanie, które zestawia ze sobą zdania nawzajem sobie nieodpowiadające. To zjawisko naturalne i poniekąd nieuniknione, z którym użytkownik InterCorp powinien się oswoić. Jest ono mniej istotne niż mała liczba poświadczeń, będąca rezultatem przeszukiwania korpusu niewielkich rozmiarów.

W tym miejscu należy odnieść się szczegółowiej do zagadnienia, jakim jest właściwe sformułowanie zapytania, które w praktyce oznacza optymalne wyszukiwanie materiału. Znaczenie prawidłowego zbudowania zapytania trudno przecenić. Wielokrotnie analiza korpusowa okazuje się nieskuteczna z powodu niewłaściwego określenia badanych jednostek i ich cech gramatycznych. Potencjalne błędy mogą być zatem następstwem przyjęcia błędnych atrybutów dla wyjściowej jednostki – translandu. Zwraca na to uwagę Milena Hebal-Jezińska w tekście *Podstawowe zasady korzystania z korpusów przy badaniu języka*²¹, w którym to znajdziemy wiele istotnych wskazówek dotyczących analizy korpusowej. Właśnie prawidłowe zbudowanie zapytania może mieć wpływ na liczbę pozyskanych poświadczeń. Ma to szczególne znaczenie w przypadku korpusu InterCorp, który dysponuje bogatą anotacją wewnętrzną i zewnętrzną²². W pracy z korpusem można skorzystać z analizy wieloczynnikowej, na przykład zawężając zakres wyszukiwania ekwiwalentów do określonej kolekcji. W przypadku polsko-rosyjskiego InterCorp, choć wyboru praktycznie nie ma (poza tekstami beletrystycznymi można skorzystać tylko

²⁰ Elektroniczny *Wielki słownik języka polskiego IJP PAN* wskazuje czasowniki *wyświadczyć*, *oddać* i *zrobić*.

²¹ M. Hebal-Jezińska, *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: W. Chlebda (red.), *Na tropach translatów...*, s. 17–30.

²² E. Kaczmarek, *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019, s. 44–46.

z tekstu Biblii i napisów filmowych), można ograniczyć na przykład wyszukiwanie wyrażień potocznych wyłącznie do kolekcji napisów filmowych. Ponadto w każdym wyszukiwaniu użytkownik ma do dyspozycji wiele atrybutów, które mogą zawęzić pole badawcze do ściśle wyselekcjonowanych tekstów. Bogata anotacja zewnętrzna umożliwia określenie podstawowych cech tekstu istotnych dla ustanowienia takich ekwiwalencji jak: przynależność do określonej części korpusu (beletrystyka [core] lub kolekcje [text.group]), typ tekstu (text.txytype), autor (text.autor), tłumacz (text.translator), miejsca wydania (text.pubplace), rok wydania (text.pubyear), tytuł tekstu (text.title), język oryginału (text.srclang), płeć autora (text.authsex), płeć tłumacza (text.transsex). InterCorp rozstrzyga kwestię zasadniczą dla badań nad ekwiwalencją, jaką jest określenie, czy tekst jest oryginałem, czy przekładem (text.original). W zależności od charakteru badań dzięki atrybutom można zawęzić pole wyszukiwawcze wyłącznie na przykład do układu tekst oryginalny — tekst przekładu. W przypadku zaawansowanych i systematycznych badań możliwe jest stworzenie własnego subkorpusu, ograniczającego przeprowadzane kwerendy do określonego typu tekstów, na przykład tekstów beletrystycznych czy administracyjnych²³. Można na przykład stworzyć subkorpus złożony z tekstów oryginalnych i ich przekładów bezpośrednich na określony język. Zbudowanie subkorpusu z wykorzystaniem określonych atrybutów jest rozwiązaniem korzystnym ze względu na uniknięcie ewentualnych błędów w ciągłym wprowadzaniu filtrów i sformułowania w konsekwencji pochopnych wniosków. Wystarczy przywołać przytoczone wyżej wyrażenie *spokojna głowa*, aby się w tym utwierdzić. Utworzenie bowiem zapytania *spokojna głowa* z zastosowaniem atrybutu *lemat* w korpusie polsko-rosyjskim nie dostarczy żadnych poświadczeń. Natomiast wyszukiwanie z użyciem wyrazów hasłowych *spokojny głowa* pozwoli uzyskać 12 poświadczeń. To wciąż mało, tym bardziej że wśród ewentualnych, sugerowanych przez InterCorp, ekwiwalentów znajdują się wyrażenia będące rezultatem błędnego automatycznego wiązania. Mimo to badacz pozyskuje materiał wyjściowy do dalszej analizy. Podobne błędy użytkowników w budowaniu zapytania przewidzieli zresztą twórcy InterCorp, usta-

²³ Sposób utworzenia subkorpusu znajdziemy w: E. Kaczmarska, A. Rosen, *Praktyczny przewodnik po Korpusie Równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowiańskich*, M. Hebal-Jeziarska (red), Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, s. 219.

nawiając od wersji 13. atrybut *lemma / word* (*lemat / forma wyrazowa*) jako domyślny.

Niemniej należy mieć na uwadze fakt, że uwzględnianie poszczególnych filtrów automatycznie pociąga za sobą zmniejszenie zakresu przeszukiwanych zasobów tekstowych, co w przypadku korpusów niewielkich rozmiarów, a do takich należy polsko-rosyjski InterCorp, może znacznie ograniczać liczbę pozyskiwanych poświadczeń, a w skrajnym przypadku uniemożliwić kwerendę.

Trudno nie odnotować w tym miejscu kluczowej roli użytkownika, któremu w świetle sformułowanych wyżej czynności z pewnością niełatwo sprostać sporym wyzwaniom. Te ostatnie wymagają bowiem od niego nie tylko odpowiednich kompetencji językowych, interkulturowych i komunikacyjnych, ale także umiejętnego wykorzystania możliwości, jakie stwarza korpus równoległy.

Aby wskazać przydatność korpusu równoległego w ustalaniu wspomnianej już ekwiwalencji przekładowej, niezbędne jest zdefiniowanie tej ostatniej, ponieważ jest ona terminem kluczowym. *Ekwiwalencja* jako termin doczekała się licznych opracowań teoretycznych we współczesnym przekładoznawstwie²⁴. Obecność korpusu równoległego w procesie ustalania par przekładowych powoduje, że najbliższe jest mi traktowanie ekwiwalencji jako konstytutywnej właściwości przekładu posiadającej charakter gradualny. Właśnie stopniowalność (zmiennność) ekwiwalencji jako jej cecha podstawowa najbardziej wyeksponowana jest w opracowaniach Wojciecha Chlebda i Romana Lewickiego. Pierwszy z nich określa typy ekwiwalentów, drugi natomiast cechy ekwiwalencji przekładowej.

Chlebda wyróżnia cztery typy ekwiwalentów:

1) ekwiwalenty pełne (adekwatne) — są ekwiwalentne wobec jednostki wyjściowej pod względem sensu i obrazowości, parametrów pragmatycznych, stylowych, a często również gramatycznych,

2) ekwiwalenty funkcjonalne — są ekwiwalentne wobec jednostki wyjściowej pod względem sensu i funkcji komunikacyjnej, różnią się obrazowością i strukturą formalną (składniową, gramatyczną),

3) ekwiwalenty zachwiane — struktury treściowe ekwiwalentów pozostają ekwiwalentne tylko częściowo wobec jednostek wyjściowych,

²⁴ M.in. J.C. Catford, *A Linguistic Theory of Translation. An Essay in Applied Linguistics*, Oxford University Press, Oxford 1965; V.N. Komissarov, *Теория перевода (лингвистические аспекты)*, Москва 1990; R. Lewicki, *Zagadnienia lingwistyki przekładu*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2017; W. Chlebda, *Ekwiwalencja i ekwiwalenty...*

4) ekwiwalenty zerowe — w języku tłumaczenia nie odnotowuje się gotowego sposobu wyrażania danej treści pochodzącej z języka oryginału²⁵.

Chyba najbardziej istotne z naszej perspektywy jest traktowanie powyższych typów ekwiwalentów jako ekwiwalentów słownikowych. Powyższa klasyfikacja uzmysławia asymetrię systemów pojęciowo-leksykalnych, w tym przypadku polszczyzny i ruszczyzny²⁶.

Lewicki wskazuje sześć cech ekwiwalencji przekładowej:

1) Ma charakter asymetryczny (jednokierunkowy). Jeden z tekstów jest oryginałem, drugi — jego przekładem. Odwrócenie tego układu nie jest możliwe.

2) Jest cechą przekładu i polega na podobieństwie przekładu do oryginału poprzez jego naśladowanie.

3) Jest stopniowalna, co oznacza, że przekład jest stopniowalny — może być w mniejszym lub większym stopniu ekwiwalentny wobec oryginału.

4) Jest względna, ponieważ może być oceniana wedle różnych kryteriów.

5) Jest hierarchiczna, ponieważ podlega jej nie tylko przekład jako całość, ale także jego poszczególne części.

6) Nie jest oparta na kryterium równoznaczności wyrażen językowych użytych w oryginale i w jego przekładzie. Jej efekt stanowi uzyskanie przekładu równoważnego w stosunku do oryginału. Zatem zarówno oryginał, jak i przekład powinny posiadać analogiczną wartość komunikacyjną dla ich odbiorców²⁷.

Choć obaj badacze koncentrują się na relacji między oryginałem a przekładem i ich wykładnikami, to jednak można dostrzec u nich dwie odmienne perspektywy. Chlebda reprezentuje podejście leksykograficzne, nastawione na notowanie par przekładowych jako jednostek języka w słowniku z całą złożonością tego procesu. U Lewickiego bardziej widać podejście tekstowe — nastawione na proces tworzenia ekwiwalentnego tekstu zarówno pod względem jego całości, jak i tworzących go elementów.

Uwzględniając powyższe, można uznać, że poszczególne przekłady sytuują się na kontinuum od najmniej do najbardziej ekwi-

²⁵ W. Chlebda, *Ekwiwalencja i ekwiwalenty...*, s. 38–40.

²⁶ Przytaczane przez autora przykłady dotyczą leksykografii przekładowej polsko-rosyjskiej.

²⁷ R. Lewicki, *Zagadnienia lingwistyki przekładu*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2017, s. 138–139.

walentnego wobec określonego oryginału. Określenia „najmniej ekwiwalentny” czy „najbardziej ekwiwalentny” to, oczywiście, pojęcia względne. Stopień ekwiwalencji – jak słusznie zauważa Lewicki – może być określany według różnych kryteriów. W literaturze przedmiotu najczęściej wskazuje się kryteria semantyczne, formalne, pragmatyczne i formalne. W zależności od autora to czy inne kryterium jest najczęściej uprzywilejowane, co często zależne jest od zakresu ekwiwalencji zachodzącej bądź między fragmentami tekstów (np. jednostkami wielowyrazowymi), bądź między całymi tekstami.

Kryterium, które przyjmuję tu za podstawę ustanawiania ekwiwalencji przekładowej, a w którym korpus równoległy może odgrywać ważną rolę, to kryterium sytuacji użycia. Właśnie ona stanowi podstawy porównania – *tertium comparationis* (dalej: TerComp). Pojęcie to w ostatnich latach na dobre zadomowiło się w badaniach nad ekwiwalencją przekładową²⁸. Świadomie pomijam tu opracowania na gruncie innych nauk czy dyscyplin, w których *tertium comparationis* służy również jako podstawa porównania. Wypada w tym miejscu przywołać jego opis pierwotny, jeszcze z lat 70. autorstwa Andrzeja Bogusławskiego, twierdzącego, że

rozstrzygającą podstawą identycznościową, czyli ich ogólnym i najwyższym *tertium comparationis* jest konkretna sytuacja wypowiedzi, znaczy to, że staramy się utworzyć taki tekst w języku przekładu, który normalny nosiciel tego języka utworzyłby w sposób maksymalnie naturalny w sytuacji właściwej tekstowi wyjściowemu²⁹.

Do zagadnienia TerComp Bogusławski powraca w monografii *Podstawy konfrontatywnej lingwistyki przekładowej*, uznając kryterium sytuacyjności za najważniejsze i zwracając uwagę na atomistyczny charakter konfrontacji przekładowej³⁰. Zdaniem badacza, nie istnieją żadne ograniczenia dotyczące wyboru określonego translatu, w szczególności ze strony języka wyjściowego, czyli języka translan-

²⁸ Zob. M. Mocarz-Kleindienst, *Tertium comparationis w badaniach nad przekładem filmowym*, „Rocznik Przekładoznawczy”, 2020, s. 233–248; W. Chlebda, *Ekwiwalencja i ekwiwalenty...*, s. 21–44; T. Piotrowski, *Tertium comparationis w przekładoznawstwie*, w: P. Stalmaszczyk (red.), *Metodologie językoznawstwa. Od genu języka do dyskursu*, Łódź, s. 175–191.

²⁹ A. Bogusławski, *Problem tertium comparationis w porównaniu lingwistycznym*, „Kwartalnik Neofilologiczny” 1976, XXIII(3), s. 295–303.

³⁰ A. Bogusławski, *Podstawy konfrontatywnej lingwistyki przekładowej*, Oficyna Wydawnicza Leksem, Łask 2013, s. 79–83.

du. Jedyne ograniczenia wskazania translatu nakłada sytuacja użycia translandu.

Analogicznie wspomnianą sytuację wypowiedzi przedstawia Piotrowski:

Zacznijmy od opisu typowej sytuacji tłumaczenia, w której występuje indywidualny akt wypowiedzenia. Ów akt niesie sąd, należy więc od niego wyjść. Analizujemy wobec tego sytuację wypowiedzenia zdania w L2, po czym ma nastąpić jego tłumaczenie na zdanie L1. Patrząc na taką sytuację czysto zewnątrznie, można wyróżnić kilka jej składowych: użytkownika języka L2, użytkownika języka L1, tłumacza między oboma, oraz kontekst, w jakim się znaleźli. Częściami tej sytuacji, kontekstu, są formy znakowe (semiotyczne), ze szczególnym uwzględnieniem form języka. Do form znakowych zaliczamy jednak także znaki niewerbalne, np. ubiór i inne oznaki statusu społecznego uczestników (tłumacz jest zwykle niejako neutralny), a także wszelki inny kontekst historyczno-społeczny, jak wiedzę o tym, co doprowadziło do zaistniałej sytuacji i co z niej może wyniknąć (np. względy dyplomatyczne, transakcje handlowe, potrzeby turysty, etc.)³¹.

Ponadto autor podkreśla, że tłumacz w całym procesie tłumaczenia powinien zachować postawę neutralną, nie wygłaszać sądów i nie przejawiać emocji. Co istotne, Piotrowski powyższe uwagi o TerComp odnosi nie tylko do procesu tłumaczenia tekstu mówionego, ale także tekstu pisanego.

W przypadku ustalania ekwiwalencji przekładowej z użyciem korpusu równoległego pojęcie TerComp nabiera specyficznego znaczenia. Jeśli przyjąć, że InterCorp zawiera w większości przekłady, to jako TerComp należałoby wskazać sytuację, w której powstaje oryginał (często angielski). To właśnie sytuacja użycia angielskiego translandu wskazuje właściwe ekwiwalenty polskie, rosyjskie czy czeskie. Mamy zatem sytuację bardziej złożoną niż ma to miejsce w klasycznym układzie tłumaczenia bezpośredniego transland – translat. Relacja naśladowania translandu przez translat nie ma miejsca; mamy natomiast zestawienie dwu translatów w większym czy mniejszym stopniu ekwiwalentnych wobec obcojęzycznego translandu. Choć zjawisko to jest niepożądane, ponieważ uobecnienie angielszczyzny czy innego języka może w pewien sposób zaburzyć prowadzoną analizę, to jednak przy dużej liczbie poświadczeń nie ma zasadniczego wpływu na wskazywane ekwiwalenty. Wskazywane zatem przez InterCorp pary przekładowe to w znakomitej większości zestawienia translatów. Deficyt tłumaczeń bezpośrednich, w których nie ujawnia się obecność angielszczyzny (głównie) jako języka pośrednika, nie jest z pewnością zaletą InterCorp.

³¹ T. Piotrowski, *Tertium comparationis w przekładoznawstwie. Metodologie językoznawstwa. Od genu języka do dyskursu*, Stalmaszczyk, Piotr (red.), Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2011, s. 175–191

Jednocześnie trzeba wyraźnie zaznaczyć, że stworzenie odpowiedniej wielkości korpusu równoległego polsko-rosyjskiego czy polsko-czeskiego, zawierającego wyłącznie tłumaczenia bezpośrednie, jest praktycznie niewykonalne. Bezsprzecznym walorem InterCorp jest uwzględnianie tłumaczeń zarówno profesjonalnych (np. literatura piękna), jak i amatorskich (np. napisy filmowe). W odniesieniu do zasobów polsko-rosyjskiego InterCorp można mówić właśnie o tych dwu sytuacjach.

Ogląd sytuacji w tekście literackim jest przecież odmienny niż w przypadku sporządzania nieprofesjonalnych napisów filmowych. W pierwszej czynność tę wykonuje profesjonalny tłumacz, w drugiej — tłumacz amator. Zróżnicowanie to z pewnością pomaga użytkownikowi korpusu podjąć trafne decyzje ekwiwalentyzacyjne.

Jak wiadomo, InterCorp zawiera teksty pisane, w zależności od zestawianych ze sobą języków mniej lub bardziej zróżnicowane gatunkowo i tematycznie. W przypadku układu polski — rosyjski wspomniane zróżnicowanie jest niewielkie. Nie licząc tekstu Biblii, wprowadzonego w wersji 10. z 2017 roku, do dyspozycji użytkownika korpusu pozostaje wspomniana literatura piękna oraz amatorskie napisy filmowe.

Wśród beletrystyki znajdziemy zarówno literaturę klasyczną, jak i współczesną. Zróżnicowanie zestawianych ze sobą przekładów jest spore. Kontext zestawia przekłady na polski i rosyjski nie tylko z angielskiego, ale także z innych języków, np. serbskiego (np. Ivo Andrić), czeskiego (np. Jaroslav Hašek, Milan Kundera), szwedzkiego (np. Astrid Lindgren), włoskiego (np. Umberto Eco), portugalskiego (np. Paulo Coelho) czy fińskiego (np. Mika Waltari). Zestawiane mogą być także polskie lub rosyjskie oryginały z przekładami na te języki (Fiodor Dostojewski, Michaił Bułhakow, Vladimir Nabokov, Boris Pasternak, Czyngiz Ajtmatow, Aleksander Puszkina, Nikołaj Ostrowskij, Ilja Ilf i Jewgienij Pietrow, Olga Tokarczuk, Stanisław Lem). Jak widać, bezpośrednie przekłady z polskiego na rosyjski czy odwrotnie stanowią jedynie uzupełnienie przekładów na polski czy rosyjski z innych języków.

Kolekcją istotną dla poszukiwania polsko-rosyjskich par przekładowych są przywołane już amatorskie napisy filmowe z platformy OpenSubtitles, która jest serwisem społecznościowym. Nie chcę tu szczegółowo omawiać cech napisów sporządzanych przez amatorów, ponieważ zrobiłem to już w artykule *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równo-*

*ległym InterCorp*³². Zastąpienie układu języków z polsko-czeskiego na polsko-rosyjski w praktyce niewiele zmienia. Autorzy najczęściej opracowują napisy w podobnych warunkach – bez specjalistycznego zaplecza technicznego. Kryteria czasowo-przestrzenne tworzenia napisów nie są przez nich ściśle przestrzegane. Niekiedy ma się wręcz wrażenie, że nie wszyscy tłumacze są świadomi ich istnienia. Ich wiedza językowa, dotycząca zarówno języka źródłowego (zwykle angielski), jak i przekładu (polski, rosyjski), jest zróżnicowana. Nie są więc zaskoczeniem liczne błędy językowe (gramatyczne, leksykalne, stylistyczne) czy tłumaczeniowe. Dla przykładu można przywołać typowe ortograficzne błędy w poświadczeniach polskich i rosyjskich:

<i>агентство</i> (zamiast <i>агентство</i>)	20 poświadczeń
<i>мороженное</i> (zamiast <i>мороженое</i>)	5 poświadczeń
<i>вообщем</i> (zamiast <i>в общем</i>)	16 poświadczeń
<i>wziąść</i> (zamiast <i>wziąć</i>)	30 poświadczeń
<i>napewno</i> (zamiast <i>na pewno</i>)	76 poświadczeń
<i>rozumieć</i> (zamiast <i>rozumiem</i>)	5 poświadczeń

Jeśli jednak porównamy liczbę przytoczonych błędów z liczbą form prawidłowych:

<i>агентство</i>	68 poświadczeń
<i>мороженое</i>	118 poświadczeń
<i>в общем</i>	342 poświadczenia
<i>wziąć</i>	509 poświadczeń
<i>na pewno</i>	1396 poświadczeń
<i>rozumiem</i>	1723 poświadczenia

to okaże się, że poświadczenia zawierające błędy, nawet te najbardziej typowe i rozpowszechnione, stanowią zdecydowaną mniejszość.

Wynika z tego, że do pozyskanych poświadczeń należy podchodzić z dużą choć nieprzesadną ostrożnością. Obraz ekwiwalencji polsko-rosyjskiej w przypadku napisów filmowych może być dodatkowo zniekształcony przez błędy, będące rezultatem automatycznego wiązania, bo takie zostało zastosowane. Niemniej błędy te pozostają w zdecydowanej mniejszości, tworząc margines zestawianych nieade-

³² A. Charciarek, *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp*, „Studia Rossica Posnaniensia” 2021, nr 1, s. 39–49.

kwatnych par przekładowych. Zalet wykorzystania napisów w ustanawianiu ekwiwalencji określonych jednostek językowych jest znacznie więcej. Dzięki tłumaczeniom sporządzanym przez autorów posiadających różne cechy społeczno-demograficzne (płeć, wykształcenie, wiek, środowisko) pozyskiwane tłumaczenia, a następnie powstające z nich pary przekładowe bywają w znacznym stopniu zróżnicowane. Nie oznacza to wcale, że zróżnicowanie to jest zjawiskiem niepożądanym; wręcz przeciwnie, dobrze obrazuje to, co rozumiemy pod pojęciem *wariantywność przekładu*. Korpus równoległy jest w tym względzie narzędziem nieocenionym, rejestrującym wybory ekwiwalentyzacyjne tłumaczy i jednocześnie demonstrującym poszukiwania optymalnego przekładu. W procesie tym — jak się wydaje — tłumacz pozostaje elementem kluczowym, dokonującym kolejnych wyborów mniej lub bardziej udanych w dużym stopniu zdeterminowanych jego wiedzą (nie tylko językową), doświadczeniem czy predyspozycjami. Jeśli uznamy — a wydaje się to oczywiste — że tekst wyjściowy może posiadać więcej niż jeden przekład, to potwierdzenie tego znajdziemy właśnie w korpusie równoległym. W niektórych przypadkach uzyskujemy wyniki idące w tysiące, niekiedy w setki, a czasami w dziesiątki poświadczeń. To pozwala na stworzenie klasyfikacji ewentualnych translatów według kryterium statystycznego, które może przesądzić o wyborze określonego ekwiwalentu.

Szczególną wartość napisy filmowe mają w odniesieniu do jednostek występujących w odmianie mówionej języka, których brak w korpusie równoległym jest szczególnie odczuwalny. Decydują o takim stanie rzeczy względy obiektywne. Jak wiadomo, teksty mówione zwykle nie są tłumaczone, a jeśli już, to rzadko. Mamy więc do czynienia poniekąd z deficytem systemowym tekstów mówionych w korpusie równoległym, który w pewien sposób napisy filmowe rekompensują.

Celem niniejszego tekstu nie było zademonstrowanie procesu ustalenia pary przekładowej za pomocą narzędzi i zasobów korpusowych, czyli wyjściem od określonych translandów i zakończeniem w postaci wskazania ich możliwych translatów. Ten ostatni etap został przeze mnie świadomie pominięty, ponieważ sam w sobie może stanowić odrębne opracowanie. Skupiłem się tu na etapie pośrednim, tj. sytuacji, gdy mamy już wyodrębnione translandy i staramy się określić zakres możliwych propozycji translatorskich, spośród których wytypowane będą translaty. Na etapie tym kluczową rolę może odegrać korpus równoległy. Co oczywiste, po spełnieniu określonych warunków. Wśród nich znajdują się te, które zależą tak od korpusu (wielkość oraz zróżnicowanie tematyczne i gatunkowe jego zasobów,

udostępnione funkcje wyszukiwawcze), jak i użytkownika (umiejętne posługiwanie się narzędziami korpusowymi, wiedza i kompetencja odnośnie do zestawianych języków). Pamiętać jednak należy, że nawet w przypadku spełnienia tych warunków użytkownik korpusu jest zobowiązany do zweryfikowania pozyskanego materiału.

Choć w porównaniu z polsko-czeskim polsko-rosyjski InterCorp korpus równoległy stwarza mniejsze możliwości w ustanawianiu par przekładowych, to może okazać się on przydatny we wskazywaniu ekwiwalentów określonych jednostek, zwłaszcza tych, które występują w języku ogólnym (beletrystyka) i odmianie potocznej (napisy filmowe). Niestety, przeprowadzenie ekwiwalentyzacji w odniesieniu do słownictwa specjalistycznego nie jest możliwe; brak tekstów o tematyce prawniczej, ekonomicznej, medycznej itp. praktycznie wyklucza pozyskanie jakichkolwiek poświadczeń.

Podsumowując, można stwierdzić, że InterCorp to narzędzie, które może skutecznie posłużyć do tłumaczenia określonych fragmentów tekstu i wskazywania możliwych ekwiwalentów. Jednocześnie korzystanie z InterCorp wymaga od użytkownika wiedzy i doświadczenia w posługiwaniu się funkcjami, które może on wykorzystywać.

Można z dużym prawdopodobieństwem założyć, że wraz z powiększaniem rozmiarów i doskonaleniem funkcji wyszukiwawczych polsko-rosyjski InterCorp będzie zyskiwał coraz więcej na znaczeniu w badaniach nad przekładem czy leksykografią przekładową. W ten sposób pozwoli on na dostarczanie danych empirycznych dotyczących istnienia bądź nieistnienia określonych jednostek języka, a w przypadku istnienia — danych statystycznych (frekwencyjnych), podnosząc rangę badań ilościowych i łącząc je z badaniami jakościowymi. Rysuje się zatem niezwykle obiecująca perspektywa dla współczesnych badań leksykograficznych i przekładoznawczych oraz językoznawczych w ogóle.

REFERENCES

- Bogusławski, Andrzej. *Podstawy konfrontatywnej lingwistyki przekładowej*. Łask: Oficyna Wydawnicza Leksem, 2013.
- Bogusławski, Andrzej. "Problem tertium comparationis w porównaniu lingwistycznym." *Kwartalnik Neofilologiczny*, 1976, z. 3: 295–303.
- Catford, John Cunnison. *A Linguistic Theory of Translation. An Essay in Applied Linguistics*. Oxford: Oxford University Press, 1965.
- Charciarek, Andrzej. "Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp." *Studia Rossica Posnaniensia*, 2021, vol. 46, no. 1: 39–49.

KORPUS RÓWNOLEGŁY A EKWIWALENTYZACJA

- Chlebda, Wojciech (ed.). *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013.
- Chlebda, Wojciech. "Ekwiwalencja i ekwiwalenty: między słownikiem a tekstami." *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2011: 21–44.
- Chlebda, Wojciech. "Korpusologia użytkowa dla początkujących i zaawansowanych." *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013: 7–15.
- Hebal-Jeziarska, Milena. „Podstawowe zasady korzystania z korpusów przy badaniu języka.” *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*. Ed. Chlebda, Wojciech, Opole: Wydawnictwo Uniwersytetu Opolskiego, 2013: 17–30.
- Kaczmarek, Elżbieta, Rosen, Alexander. „Praktyczny przewodnik po korpusie równoległym InterCorp.” *Praktyczny przewodnik po korpusach języków słowiańskich*. Ed. Hebal-Jeziarska, Milena. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 207–231.
- Kaczmarek, Elżbieta. *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2019.
- Komissarov, Vilen Naumovich. *Teoriya perevoda (lingvisticheskiye aspekty)*. Moskwa: Vysshaya shkola, 1990 [Комиссаров, Вилем Наумович. *Теория перевода (лингвистические аспекты)*. Москва: Высшая школа, 1990].
- Lewandowska-Tomaszczyk, Barbara. Ed. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Lewicki, Roman. *Zagadnienia lingwistyki przekładu*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2017.
- Moczarz-Kleindienst, Maria. "Tertium comparationis w badaniach nad przekładem filmowym." *Rocznik Przekładoznawczy*, 2020: 233–248.
- Piotrowski, Tadeusz. "Tertium comparationis w przekładoznawstwie." *Metodologie językoznawstwa. Od genu języka do dyskursu*. Ed. Stalmaszczyk, Piotr. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2011: 175–191.
- Tabakowska, Elżbieta. *Gramatyka i obrazowanie. Wprowadzenie do językoznawstwa kognitywnego*. Kraków: Wydawnictwo Oddziału PAN, 1995.

ONLINE SOURCES

- Bańczyk, Lucja, Dybalska, Renata, Vavřín, Martin. Korpus InterCorp — polština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- Rajnochová, Natálie, Runštuková, Naděžda, Vavřín, Martin. Korpus InterCorp — ruština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- Rosen, Alexandr, Vavřín, Martin, Zasina, Adrian Jan. Korpus InterCorp — čeština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020 <<http://www.korpus.cz>> (10.12.2021).
- <<https://ruscorpora.ru>>.
- <<http://pol-ros.polon.uw.edu.pl>>.