

ANDRZEJ CHARCIAREK

Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp

Subtitles as the source of translation equivalents in the Polish-Czech parallel corpus InterCorp

Abstract. The article focuses on using film subtitles in the process of establishing Polish-Czech language pairs. Research material comes from the resources of the OpenSubtitles website with non-professional subtitling. The research material was analysed using KonText – the main search tool in InterCorp. As a starting point, the author analysed a dictionary entry for the Polish exclamation *cholera* in a traditional Polish translation dictionary like *Polish-Czech Dictionary* (Oliva). It presents translation equivalents in translation pairs such as *cholera – sakra, hergot, do cholery – do hajzlu, idź do cholery – táhni ke všem čertům, do jasnej cholery – k sakru, do hajzlu, cholera mnie bierze – čerti mě berou*. All the translation equivalent pairs have been analysed using the corpus analysis with three functionalities available in the KonText search engine: type of question (*typ dotazu*), content of the equivalent (*obsahuje*), and limiting search (*omezit hledání*). The conducted corpus analysis allowed for the verification of information in the traditional dictionary, but also to find new equivalents that have not been registered yet which could also be used in translation.

Keywords: OpenSubtitles, InterCorp, translation equivalence, Czech, Polish

Andrzej Charciarek, Uniwersytet Śląski w Katowicach, Katowice – Polska, andrzej.charciarek@us.edu.pl, ORCID ID: <https://orcid.org/0000-0001-5651-5730>

Prace teoretyczne poświęcone leksykografii to opracowania dotyczące głównie leksykografii jednojęzycznej. Na niej koncentrowali się przede wszystkim badacze, zarówno leksykografowie teoretycy, jak i słownikarze praktycy. To, że leksykografia jest integralną częścią językoznawstwa, co wydaje się obecnie oczywiste, nie zawsze było powszechne. Tadeusz Piotrowski uznawał leksykografię, rozumianą jako sztukę tworzenia słowników, za dziedzinę autonomiczną, która w znacznej mierze korzysta z metod analizy językoznawczej, ale jest częścią kultury humanistycznej (Piotrowski 1994: 10). Tę odrębność leksykografii jeszcze bardziej eksponuje Mirosław Bańko w tytule swojej monografii *Z pogranicza leksykografii i językoznawstwa* (Bańko). Paradoksalnie jednak jest ona w całości poświęcona

kwestiom językowym wskazującym na ścisły związek językoznawstwa i leksykografii. Według Piotra Żmigrodzkiego ten związek jest bezsporny. Postrzega on współczesną leksykografię jako integralną część językoznawstwa, a słownik jako wynik analizy lingwistycznej poszerzonej o wiedzę z innych dziedzin, wśród których wskazuje fleksję, składnię, semantykę, kombinatorykę leksykalną, stylistykę czy pragmatykę (Żmigrodzki 21–22).

Należy jednak mieć na uwadze fakt, że perspektywy widzenia problematyki leksykograficznej wspomnianych autorów są odmienne. Piotr Żmigrodzki i Mirosław Bańko to leksykografowie polscy, jednojęzyczni, Tadeusz Piotrowski jest anglistą, przekładoznawcą skupionym głównie na zagadnieniach leksykografii przekładowej. Stąd różne spojrzenia na wiele kwestii dla budowy słownika podstawowych, a dotyczących metaleksykografii językoznawczej, czyli procedur opracowywania słowników. Te ostatnie odnoszą się do rozstrzygnięć dla leksykografii fundamentalnych, tj. makrostruktury i mikrostruktury słownika, bo te, z oczywistych względów, będą dla leksykografii jednojęzycznej i przekładowej różne.

Jak wiadomo, podstawową czynnością dla leksykografa jest zgromadzenie korpusu tekstów. Metoda tradycyjna stosowana w epoce przedkomputerowej sprowadzała się do stworzenia kartoteki, w której notowane były jednostki jedno- lub wielowyrazowe w określonym kontekście. Źródłem najczęściej były teksty pisane (np. beletrystyka), rzadziej mówione (np. zapis wypowiedzi). One to, po określeniu siatki haseł, stanowiły podstawę struktury artykułu hasłowego.

Powstanie narodowych korpusów lingwistycznych diametralnie zmieniło sytuację, stworzyło nowe możliwości w aspekcie pozyskiwania, gromadzenia, segregowania i analizy materiału językowego. Zautomatyzowanie w znacznym stopniu pracy leksykografa w połączeniu z wielkimi zbiorami tekstów, w dodatku tworzonymi z uwzględnieniem kryterium reprezentatywności, zwiększyło nie tylko szybkość przeprowadzenia analizy, ale także jej wiarygodność. Warto jednak w tym miejscu podkreślić, że bez względu na nowe możliwości istota pracy leksykografa się nie zmieniła – to on w dalszym ciągu wskazuje materiał językowy, precyzuje siatkę haseł i przyjmuje określoną koncepcję opisu leksykograficznego, a przede wszystkim samodzielnie dokonuje analizy wyekscerpowanego materiału.

Z punktu widzenia leksykografa przekładowego dobór źródeł tekstowych wydaje się kluczowy i nic dziwnego, że o nim w kontekście pozyskiwania ekwiwalentów przekładowych mówiło się praktycznie od zawsze. Bolączką podstawową tradycyjnych słowników przekładowych było ich zorientowanie na system, czyli zestawianie ze sobą jednostek systemu bez szczególnego uwzględniania kontekstu ich użycia. Było to wynikiem strukturalistycznego podejścia do języka, które skutecznie zdominowało poszczególne leksykografie narodowe w XX

wieku. Piszę „narodowe”, mając na względzie opracowania jednojęzyczne, czyli wielkie słowniki poszczególnych języków. Słowniki przekładowe były niejako produktem wtórnym, w dużym stopniu implementującym, niekiedy nieudolnie, rozwiązania stosowane w leksykografii jednojęzycznej. Analogicznie autorzy słowników przekładowych orientowali się zwykle na źródła systemowe, w których wyodrębniali, nierzadko w sposób niejasny, jednostki języka, a następnie, niekiedy czysto intuicyjnie, ustanawiali ekwiwalencję między jednostkami języka wyjściowego a jednostkami języka przekładu.

W ostatnich dwóch-trzech dekadach ten stan rzeczy zaczął się zmieniać, a to za sprawą dostrzeżenia przez leksykografów oczekiwań tłumaczy odnośnie do tego, co słownik przekładowy powinien zawierać. A powinien notować ekwiwalenty przekładowe, bo przecież sporządzając tłumaczenie tekstu, tłumacz sięga do słownika właśnie po nie. Jeśli chodzi o jednostki jednowyrazowe, to sprawa wydaje się prosta, a przynajmniej prostsza. W takim czy innym słowniku przekładowym prędzej czy później się one znajdują. Problem jednak pojawia się w procesie notowania jednostek wielowyrazowych i dotyczy dwu jego etapów. Pierwszy sprowadza się do umiejętnego wyodrębnienia wielowyrazowca w języku wyjściowym, drugi – do trafnego wskazania jego ekwiwalentu czy ekwiwalentów w języku przekładu. Podkreślam tę dwuetapowość szczególnie, ponieważ nieadekwatne pary przekładowe są nierzadko wynikiem nietrafnego czy niepełnego określenia znaczenia metaforycznego jednostki wyjściowej. Błąd czy nieścisłość w określeniu znaczenia translandu (jednostki języka źródłowego) automatycznie skutkuje wskazaniem nieadekwatnego translatu (jednostki języka przekładu). Zakładając nawet doskonałą kompetencję językową leksykografa, będzie on zawsze rodzimym użytkownikiem jednego języka. Rzecz jasna, możemy sobie wyobrazić leksykografa będącego osobą bilingwalną, ale przyznajmy, że to wyjątek. Można się tylko zastanawiać, czy lepiej, aby rodzimy użytkownik języka ujawnił się po lewej czy prawej stronie słownika przekładowego.

W tytule niniejszego artykułu wskazane są napisy filmowe jako źródło ekwiwalentów w przekładzie polsko-czeskim w korpusie równoległym InterCorp bez sprecyzowania źródła tych napisów, którym jest platforma OpenSubtitles. To oznacza, że InterCorp oferuje amatorskie napisy filmowe, tworzone przede wszystkim przez miłośników bądź kina, bądź języków. Autorami są zazwyczaj ludzie młodzi, choć nie tylko. Można znaleźć tłumaczenia, których twórcy wyraźnie zdradzają swoje niemałe zdolności językowe i translatorskie. Naturalnie, zdarzają się tłumaczenia złe, poświadczające słabą znajomość języka czy języków ich autorów.

Każdy, kto korzystał z korpusu InterCorp, wie o tym, że służy on przede wszystkim do badania języka czeskiego. Wszystkie inne języki z nim zestawiane służą dobrze poszczególnym leksykografom przekładowym, np. czesko-polskiej,

czesko-angielskiej czy czesko-rosyjskiej. Mierząc przydatność korpusu równoległego jego zasobami, InterCorp najczęściej ma do zaoferowania czesko-angielskiej leksykografii przekładowej. Czesko-angielska część jest największym dwujęzycznym korpusem równoległym dostępnym w InterCorp.

Dwunasta wersja polsko-czeskiego InterCorp sytuuje się na szóstym miejscu pod względem wielkości wśród możliwych do zestawienia korpusów równoległych. Decyduje o tym wielkość polskiego modułu tekstów – ponad 87 mln wyrazów. Na tę liczbę składają się teksty literatury pięknej (ponad 25 mln wyrazów), teksty prawne z korpusu Acquis Communautaire (blisko 20 mln wyrazów), sprawozdania z obrad Parlamentu Europejskiego (prawie 13 mln), teksty publicystyczne i wiadomości ze stron internetowych VoxEurop (niespełna 2,5 mln wyrazów), amatorskie napisy filmowe z bazy OpenSubtitles (prawie 27 mln wyrazów) oraz przekłady Biblii (ponad pół miliona wyrazów).

Trzeba mieć świadomość, iż zestawiane z sobą teksty polskie i czeskie to zwykle nie układ oryginał – przekład, a układ przekład – przekład. Najczęstszym pośrednikiem, tj. oryginałem, dla przekładu polskiego i czeskiego jest tekst po angielsku. Niestety, przekłady bezpośrednie polsko-czeskie czy czesko-polskie stanowią tylko znikomą część tekstów korpusu równoległego. W przypadku napisów filmowych, które tu omówię, przekłady bezpośrednie polsko-czeskie to już zupełny margines, są praktycznie nieobecne. O tym, z jakimi tekstami ma się do czynienia, użytkownik InterCorp może się dowiedzieć dzięki szczegółowej anotacji zewnętrznej tekstów. W przypadku nieprofesjonalnych napisów filmowych z platformy OpenSubtitles podawana informacja jest skąpa – wiemy, jaki jest język oryginału i kierunek tłumaczenia. Otrzymujemy więc spory zasób materiału językowego, praktycznie niezweryfikowanego i pochodzącego od autorów, których kwalifikacji translatorskich i kompetencji językowych nie jesteśmy w stanie w żaden sposób sprawdzić. Mało tego, dysponujemy napisami filmowymi, które w odróżnieniu od napisów profesjonalnych nie powstały zgodnie z określonymi kryteriami, z zastosowaniem specjalistycznego sprzętu i oprogramowania. Wspomnę tu tylko o liście dialogowej, którą dysponuje profesjonalny autor napisów filmowych, uwzględniając przy ich tworzeniu różnego rodzaju ograniczenia natury technicznej (czasowe, przestrzenne, sekwencyjne, graficzne). Profesjonalny sprzęt, którym dysponuje, pozwala mu na dokonywanie wielu operacji (zwykle kondensacyjnych, adaptacyjnych) mających wpływ na ostateczny kształt napisów. Jego kompetencja językowa pozwala nie tylko na uniknięcie błędów, ale przede wszystkim na umiejętną ingerencję w tekst tłumaczenia, aby był on dostosowany do wymogów technicznych. Znamienne jest, jak zauważa Teresa Tomasziewicz, że dialogi nierzadko są wspomagane przez obraz, mimikę i gestykulację (Tomaszkiewicz 116). Wszystkie te czynniki mają wpływ na ostateczną formę napisów na ekranie.

Amatorskie napisy filmowe nie powstają w specjalistycznym studiu, ale w zaciszu domowym, w którym brakuje wyżej wymienionego zaplecza technicznego i listy dialogowej. Sami autorzy napisów zamieszczanych na platformie OpenSubtitles piszą, że zwykle opracowują je na podstawie odsłuchu ścieżki dźwiękowej filmu, a kryteriów czasowo-przestrzennych przestrzegają nierygorystycznie. Wydaje się, że większość autorów napisów nieprofesjonalnych nawet nie wie o istnieniu cech, którymi odznaczają się napisy profesjonalne, tj. czytelności, przyswajalności, dyskretności i naturalności (Belczyk 11). Nie stosują, a w każdym razie stosują rzadziej, co również można oceniać pozytywnie, różnego rodzaju zabiegów upraszczających język oryginału po to, aby był lepiej przyswajalny dla widza. Ponadto, co także przyznają, wykorzystują swoją wiedzę językową, której poziom odpowiada wiedzy przeciętnego użytkownika języka. Potwierdzają to poświadczenia pozyskiwane z platformy OpenSubtitles, dokumentujące nie tylko błędy tłumaczeniowe, ale i językowe (ortograficzne również). Nie sposób jednak w tym rzeczywiście niedoskonałym czasem materiale nie dostrzec walorów.

Od dawna podstawowym problemem korpusów językowych był deficyt zawartych w nich tekstów mówionych. Co prawda napisów filmowych nie można traktować jako wiernych zapisów języka mówionego, to jednak mają one wiele podobieństw. Najczęściej odzwierciedlają wypowiedzi spontaniczne w sytuacji nieoficjalnej. Tych właśnie często brakuje w korpusach narodowych. Przydatność tego typu zapisów z pewnością docenią użytkownicy korzystający z niewielkiego podkorpusu języka mówionego Narodowego Korpusu Języka Polskiego. Znajdziemy zatem w napisach filmowych połączenia o dużym stopniu idiomatyczności, oczywiście dla rodzimych użytkowników określonego języka, ale już nie dla nierodzimych. Będą to przede wszystkim jednostki, które możemy nazywać komunikacyjnymi, często nierejestrowane w słownikach przekładowych. Nie są one ani frazeologizmami w tradycyjnym znaczeniu, ani przysłowiami itp. Jeśliby takowymi były, z pewnością zostałyby zanotowane w odpowiednich słownikach specjalistycznych. Ponieważ nie stanowią tzw. twardej frazeologii, często pozostają poza polem widzenia leksykografów. To, co jest ewidentnym mankamentem napisów filmowych, czyli nieprzestrzeganie kryteriów tworzenia napisów, może paradoksalnie okazać się pozytywnym w analizie leksykograficznej. W świadomości wielu tłumaczy amatorów najważniejszą zasadą jest dochowanie w tłumaczeniu wierności oryginałowi nawet za cenę nierespektowania ograniczeń czasowo-przestrzennych. Ponadto mamy w przypadku amatorskich napisów filmowych do czynienia ze swoistym wielogłosem społeczeństwa. Są one opracowywane przez osoby w różnym wieku, różnej płci, mające różne wykształcenie, pochodzące z różnych środowisk, reprezentujące różne światopoglądy. Wszystkie te cechy sprawiają, że w napisach filmowych ujawniają się cechy idiolektalne charakterystyczne dla mowy ich autorów. Mogą one dotyczyć

różnych poziomów ich języka, np. leksykalnego, składniowego, semantycznego czy pragmatycznego. To zróżnicowanie autorów jest niewątpliwą zaletą napisów w korpusie równoległym. Te same odcinki tekstu źródłowego mogą być i są tłumaczone w różny sposób, a wiele tych tłumaczeń można uznać za adekwatne. Jeśli przyjąć, że ekwiwalencja ma charakter gradualny, to każde tłumaczenie może być oceniane według większego lub mniejszego stopnia ekwiwalencji rozumianej jako podobieństwo do oryginału. Oczywiście jest, że wspomniane podobieństwo oryginału i jego tłumaczenia musi dotyczyć określonych ich cech. Jednak bez względu na ich wskazanie należy pamiętać, że „ostatecznym celem ekwiwalencji przekładowej jest zapewnienie, aby [przekład – A.C.] zyskał [...] dla swych odbiorców analogiczną wartość komunikacyjną w stosunku do wartości komunikacyjnej, jaką oryginał ma dla swoich odbiorców” (Lewicki 139).

W przypadku napisów filmowych z platformy OpenSubtitles mamy sytuację nietypową, a mianowicie porównujemy dwa przekłady, dla których oryginałem jest tekst angielski. Angielszczyzna jest językiem pośrednikiem, co z pewnością nie ułatwia wskazania pary ekwiwalentów przekładowych. Mamy natomiast duże prawdopodobieństwo adekwatności tłumaczeń angielsko-czeskich i angielsko-polskich, które w konsekwencji mogą ułatwić wskazanie ekwiwalencji określonych jednostek polskich i czeskich. Należy w tym miejscu podkreślić również pewną wyjątkowość napisów filmowych będących tłumaczeniem uwzględniającym w dużym stopniu elementy pragmatyczne. Tłumacz praktycznie ma do dyspozycji wszystkie elementy sytuacji komunikacyjnej, a przede wszystkim charakterystyki rozmówców, ich intencje i cele, miejsce i czas kontaktu. Innymi słowy, tłumacz opracowujący napisy ma zadanie w pewien sposób ułatwione, ponieważ obserwuje funkcjonowanie oryginału w określonym kontekście, w którym to musi umieścić tekst tłumaczenia mający równoważną wartość komunikacyjną. Napisy filmowe, a następnie korpus równoległy, który je zawiera, rejestrują określoną typowość i powtarzalność zachowań werbalnych w określonych sytuacjach, np. wzburzenia, dezaprobaty, irytacji, zachwyty itp. I choć mamy w ich przypadku do czynienia z materiałem polskim i czeskim, które w różnym stopniu naśladują oryginał (zwykle angielski), to można założyć, że w większym lub mniejszym stopniu są one wobec niego ekwiwalentne. W każdym razie nie można twierdzić, że związku między przekładami nie ma. Translady (jednostki języka przekładu, np. polski i czeski) są determinowane przez ten sam obcojęzyczny transland (jednostkę języka źródłowego – angielski), który – cokolwiek by mówić – w pewien sposób uczestniczy w procesie ustanawiania pary przekładowej. Niemniej jednak rezygnacja z przekładów z innych języków i badanie wyłącznie polsko-czeskich tłumaczeń bezpośrednich spowodowałaby drastyczne zmniejszenie liczby dokumentów, a co za tym idzie, także liczby poświadczeń. Wiadomo, że zbyt mała liczba poświadczeń może uniemożliwić przeprowadzenie wiarygodnych, zobiektywizowanych badań.

Przyjmując zatem amatorskie napisy z platformy OpenSubtitles za materiał egzemplifikacyjny, należy mieć świadomość zarówno jego zalet, jak i wad. Do tych ostatnich należy zaliczyć także automatyczne wyrównanie tekstów (ang. *alignment*), które nie gwarantuje adekwatności zestawianych ze sobą polskich i czeskich ekwiwalentów. Z tej przyczyny błędnych poświadczeń dostarczy InterCorp praktycznie zawsze i trzeba być na nie przygotowanym, podobnie jak na wskazywane w większości trafne odpowiedniki. Co prawda oprogramowanie do automatycznego opracowania tekstów w InterCorp jest konsekwentnie udoskonalane, ale w odróżnieniu od metody ręcznej nieraz zawodzi. Na wskazywanie przypadkowych jednostek, znajdujących się w bezpośrednim sąsiedztwie, a nie właściwych ekwiwalentów należy być przygotowanym.

Na wstępie określę to, co będzie obiektem krótkiej analizy, czyli jedno z wyrażeń ekspresywnych. W ujęciu Stanisława Grabiasa to: „te znaki semiotyczne, za pomocą których nadawca wyraża swój stosunek do otaczającej go rzeczywistości lub w których niezależnie od intencji nadawcy przejawiają się cechy jego osobowości. Za pomocą znaków emocjonalnych zaś nadawca dokonuje negatywnej lub pozytywnej oceny zjawisk” (Grabias 37).

Wyjściową jednostką ustanowimy wykrzyknik *cholera*, definiowany w elektronicznym *Wielkim słowniku języka polskiego* (źródło elektroniczne) jako przekleństwo z kwalifikacją tematyczną „zachowania i wyrażenia nieakceptowane społecznie”¹. Słownik ten podaje również tworzone z jego udziałem frazeologizmy: *cholera bierze kogoś*, *cholera jasna*, *do cholery*, *za cholere*. Poszukiwania jego czeskiego ekwiwalentu w badanej funkcji w słowniku polsko-czeskim pod redakcją Karela Olivy przynoszą następujące pary przekładowe:

cholera – sakra, hergot
do cholery – do hajzlu
idź do cholery – táhni ke všem čertům
do jasnej cholery – k sakru, do hajzlu
cholera mnie bierze – čerti mě berou (Oliva 129)

Wyszukiwarka KonText w korpusie równoległym InterCorp zawężonym do kolekcji napisów filmowych przynosi następujące wyniki dla:

cholera
cholera – sakra (3853 poświadczenia)
cholera – do prdele (1884 poświadczenia)

¹ https://www.wsjp.pl/do_druku.php?id_hasla=44488&id_znaczenia=4694881. Web. 18.06.2020.

cholera – kurva (1122 poświadczenia)
cholera – zatraceně (800 poświadczeń)
cholera – do hajzlu (652 poświadczenia)
cholera – kruci (468 poświadczeń)
cholera – hergot (150 poświadczeń)
cholera – krucinál (122 poświadczenia)
cholera – k sakru (32 poświadczenia)

cholera jasna

cholera jasna – sakra (89 poświadczeń)
cholera jasna – do prdele (64 poświadczenia)
cholera jasna – zatraceně (55 poświadczeń)
cholera jasna – kurva (28 poświadczeń)
cholera jasna – do hajzlu (20 poświadczeń)
cholera jasna – kruci (15 poświadczeń)
cholera jasna – krucinál (13 poświadczeń)
cholera jasna – k sakru (4 poświadczenia)
cholera jasna – hergot (2 poświadczenia)

do cholery

do cholery – sakra (1509 poświadczeń)
do cholery – kurva (365 poświadczeń)
do cholery – zatraceně (214 poświadczeń)
do cholery – do prdele (133 poświadczenia)
do cholery – k sakru (71 poświadczeń)
do cholery – kruci (67 poświadczeń)
do cholery – krucinál (64 poświadczenia)
do cholery – hergot (59 poświadczeń)
do cholery – do hajzlu (39 poświadczeń)

do jasnej cholery

do jasnej cholery – sakra (30 poświadczeń)
do jasnej cholery – zatraceně (15 poświadczeń)
do jasnej cholery – kurva (11 poświadczeń)
do jasnej cholery – do prdele (10 poświadczeń)
do jasnej cholery – hergot (6 poświadczeń)
do jasnej cholery – krucinál (5 poświadczeń)
do jasnej cholery – k sakru (2 poświadczenia)
do jasnej cholery – do hajzlu (1 poświadczenie)
do jasnej cholery – kruci (1 poświadczenie)

Przytoczone dane korpusowe nie są ściśle i nie mogą być uważane za jedyną podstawę ustanowienia par przekładowych. Wśród pozyskanych poświadczeń znajduje się sporo zdublowanych, wiele z nich się pokrywa, np. w liczbie poświadczeń *cholera – sakra* uwzględnione są również poświadczenia pary *cholera jasna – sakra*.

Dzięki funkcjom wyszukiwarki KonText można doprecyzować wiele kwerend korpusowych. Nie na wszystkie operacje pozwalają jej funkcje, ale na niektóre – tak. W przeprowadzonej analizie wykorzystane zostały: rodzaj zapytania (*typ dotazu*), zawartość ekwiwalentu (*obsahuje*), zawężić wyszukiwanie (*omezit hledání*). Bez względu na niedoskonałości natury technicznej, a także zwykłe błędy tłumaczeniowe i językowe² autorów napisów zgromadzony materiał jest cenny sam w sobie. Oczywiście pod warunkiem traktowania go jako materiał wyjściowy do dalszego opracowania.

Jak widać, pozyskane poświadczenia w dużym stopniu potwierdzają trafność analizy leksykograficznej w słowniku polsko-czeskim (Oliva). Niemniej wiele kwestii, czego można było oczekiwać, korpus równoległy pomaga uściślić. Jeśli pozycja pierwsza ekwiwalentu *sakra* nie ulega wątpliwości, to można się zastanawiać nad zasadnością podawania na drugim miejscu ekwiwalentu *hergot*, który jest germanizmem (czes. *panebože*) i tworzy parę przekładową bardziej z semantycznie bliskim *můj Bože*. Nie zostały zanotowane w słowniku papierowym ekwiwalenty *kurva* i *zatraceně*, które zajmują pod względem poświadczeń trzecie i czwarte miejsce. Zwróćmy uwagę, że korpus podaje zarówno parę przekładową *cholera – do prdele*, jak i *do cholery – do prdele*. Ta pierwsza nie jest notowana w odnośnym słowniku przekładowym, druga – tak. Wydaje się, że jest to przykład często stosowanej w leksykografii zasady, aby maksymalnie zbliżyć pod względem formalnym elementy tworzące parę przekładową, co nierzadko jest rozwiązaniem błędnym. Podobnie częściej wyszukiwarka KonText wskazuje parę ekwiwalentów *cholera – do hajzlu* niż, zdawałoby się, bliższą sobie (bo z przyimkami) *do cholery – do hajzlu*. To pokazuje, że zasada *per analogiam* nie zawsze naprowadza na adekwatne rozwiązanie ekwiwalentyzacyjne. Znamienna jest duża liczba poświadczeń wykrzyknika *kruci* (jego wariant *krucinál* używany jest rzadziej), który również nie jest notowany we wspomnianym słowniku.

² W napisach z platformy OpenSubtitles znajdziemy sporo błędów ortograficznych, np. część czesko-polska rejestruje 1206 poświadczeń *doprdele* i 724 poświadczenia *ksakru*, w części polsko-czeskiej znajdziemy 9 poświadczeń *holera*, 5 – *do holery*. Znamienne jest, że jeśli polskie poświadczenia można oceniać jako oczywiste błędy ortograficzne, to czeskie już niekoniecznie. Są one często traktowane jako wyrażenia ekspresywne przynależne do czeszczyzny potocznej nieliterackiej, choć *Slovník nespisovné češtiny* (Hugo) ich nie notuje.

Odrębnego omówienia wymagają przytoczone w słowniku *idź do cholery – táhnike všem čertům* i *cholera mnie bierze – čerti mě berou*. Mankamentem ich analizy korpusowej jest niewielka liczba poświadczeń, co nie pozwala na sformułowanie wiarygodnych wniosków. Natomiast pewne spostrzeżenia poczynić można. Można sądzić, że frazem *ke všem čertům* bez czasowników *táhnout, poslat* i *jít* może być ekwiwalentem *do cholery*, przynajmniej nieliczne poświadczenia to potwierdzają. Z formami rozkaźników *táhni* i *jdi* frazem ten stanowi bardziej ekwiwalent *idź do diabła* niż *idź do cholery*. Zweryfikowanie frazematów *cholera mnie bierze – čerti mě berou* za pomocą korpusu napisów filmowych nie jest możliwe, nie ma ani jednego takiego poświadczenia. W całym korpusie InterCorp znajdziemy tylko jedno poświadczenie *čerti mě berou – diabli mnie biorą* z powieści Witolda Gombrowicza *Trans-Atlantyk*. To dowodzi, że oba nie są w aktywnym użyciu Czechów i Polaków.

Pora na sformułowanie wniosków przeprowadzonej analizy korpusowej, dla której punktem wyjścia był już trochę leciwy (chodzi o czas jego powstania) tradycyjny słownik polsko-czeski Karela Olivy (Oliva). Napisy filmowe, będące jedną z kolekcji InterCorp, nie podają gotowych rozwiązań przekładowych, podpowiadają jedynie możliwości tłumaczenia określonych fragmentów tekstu, w tym przypadku polskiego wykrzyknika *cholera* wraz z wariantami w języku czeskim. W wielu przypadkach pozyskane dane korpusowe pozwoliły na uszczegółowienie, a przede wszystkim poszerzenie informacji zawartych w tradycyjnym słowniku przekładowym. Napisy filmowe opracowane przez osoby zróżnicowane pod względem cech socjodemograficznych pozwoliły na wskazanie ekwiwalentów dotąd niedostrzeżonych. Poświadczenia idące często w setki, a nawet tysiące wskazały jednostki najczęściej używane przez rodzimych użytkowników języka polskiego i czeskiego. Okazało się, co w pewien sposób było do przewidzenia, że wyrażenia ekspresywne mają z reguły więcej ekwiwalentów przekładowych. Jednym z kryteriów pewnej ich hierarchii mogą być dane korpusowe – pokazują te częściej i rzadziej używane, choć mające tę samą wartość komunikacyjną. Rzecz jasna, analiza ilościowa w ustaleniu pary przekładowej pomaga, ale nie stanowi wyłącznego kryterium. Warto jednak pamiętać o tym, że dane frekwencyjne dokumentują ważną cechę jednostek języka – ich odtwarzalność.

Należy podkreślić, że mimo materiału językowego wielce niedoskonałego, z tak zwanej drugiej ręki (przekłady z angielskiego na czeski i polski), można na podstawie jego analizy porządkować pary przekładowe, wskazując adekwatność jednych, podważając ekwiwalencję drugich. W pewnych przypadkach korpus równoległy okazuje się bezradny. Przeprowadzona analiza potwierdza przydatność narzędzi i zasobów korpusowych w ustalaniu ekwiwalentów przekładowych. W odróżnieniu od wyników analizy materiału korpusowego artykuł hasłowy *cholera* w tradycyjnym słowniku przekładowym nie rejestruje

wielu ewentualnych ekwiwalentów. Przyczyny tego stanu rzeczy są dwie: po pierwsze, brak możliwości skorzystania w momencie tworzenia wspomnianego słownika z zasobów i instrumentarium korpusowego, a po drugie, rejestrowanie języka polskiego i czeskiego z kartotek sprzed kilkudziesięciu lat. Reasumując, bez względu na to, że pozyskany materiał językowy jest dalece niedoskonały, to jednak skutecznie pozwala zrewidować wybory nietrafne, uszeregować warianty według kryterium frekwencyjnego i, co chyba najważniejsze, wskazać nowe ekwiwalenty przekładowe.

Bibliografia

- Bańko, Mirosław. *Z pogranicza leksykografii i językoznawstwa*. Warszawa, Wydział Polonistyki Uniwersytetu Warszawskiego, 2001.
- Belczyk, Arkadiusz. *Tłumaczenie filmów*. Wilkowice, Wydawnictwo „Dla szkoły”, 2007.
- Hugo, Jan, red. *Slovník nespisovné češtiny*. Praha, Maxdorf, 2009.
- Grabias, Stanisław. *O ekspresywności języka. Ekspresja a słowotwórstwo*. Lublin, Wydawnictwo Lubelskie, 1981.
- Lewicki, Roman. *Zagadnienia lingwistyki przekładu*. Lublin, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2017.
- Oliva, Karel, red. *Polsko-český slovník*. Praha, Academia, 1999.
- Piotrowski, Tadeusz. *Z zagadnień leksykografii*. Warszawa, Wydawnictwo Naukowe PWN, 1994.
- Tomaszkiewicz, Teresa. *Przekład audiowizualny*. Warszawa, Wydawnictwo Naukowe PWN, 2006.
- Wielki słownik języka polskiego*. Web. 18.06.2020. <https://www.wsjp.pl>.
- Żmigrodzki, Piotr. *Słowo – słownik – rzeczywistość. Z problemów leksykografii i metaleksykografii*. Kraków, Lexis, 2008.

Korpusy

- Bańczyk Łucja, Dybalska Renata, Vavřín Martin. Korpus InterCorp – polština, verze 12 z 12.12.2019. Ústav Českého národního korpusu FF UK, Praha 2019. Web. 20.01.2020. <http://www.korpus.cz>.
- Rosen Aleksander, Vavřín Martin, Zasina Adrian Jan. Korpus InterCorp – čeština, verze 12 z 12.12.2019. Ústav Českého národního korpusu FF UK, Praha, 2019. Web. 20.01.2020. <http://www.korpus.cz>.