

ELŻBIETA KACZMARSKA
Instytut Sławistyki Zachodniej i Południowej
Wydział Polonistyki
Uniwersytet Warszawski
ul. Krakowskie Przedmieście 26/28
00-927 Warszawa
tel. (+48) 22 55 20 311
e-mail: e.h.kaczmarska@uw.edu.pl

ALEXANDR ROSEN
Ústav teoretické a počítačnické lingvistiky
Univerzita Karlova, Filozofická fakulta,
U kříže 8, 150 00 Praha 5
tel. +420 221 619 752
e-mail: Alexandr.Rosen@ff.cuni.cz

ANALIZA KORPUSOWA CZESKIEGO CZASOWNIKA *POSTRÁDAT* I JEGO POLSKICH EKWIWALENTÓW

SŁOWA KLUCZOWE: Psych verbs, korpus równoległy, język czeski, język polski, ekwiwaleńt leksykalny.

KEYWORDS: Psych verbs, parallel corpus, Czech, Polish, lexical equivalent.

A CORPUS-BASED ANALYSIS OF THE CZECH VERB *POSTRÁDAT* AND ITS POLISH EQUIVALENTS

ABSTRACT: Polysemous verbs may be difficult to translate even into a closely related language. A structurally similar equivalent is often unavailable or suboptimal due to the different patterning of a given semantic field in the target language. Czech psych verbs translated into Polish represent a case in point. Following previous studies of a number of such verbs, the focus of the contribution is on *postrádat* ‘miss’. Lexical entries in several monolingual, bilingual, and valency dictionaries are contrasted with concordances found in a parallel corpus, offering a colorful picture of various strategies, many of them structurally different from the original. Despite relatively low frequencies, some preferences in the choice of equivalents have been detected, namely for *brak* or *brakować*.

Interestingly, in Polish originals the equivalents of the Czech target verb *postrádat* are often preceded by *bez* ‘without’ or *bez-*, the homograph prefix. The corpus offers more equivalents if neither Czech nor Polish text is the original. However, the results are more difficult to interpret due to the unclear status of the translation process.

Komunikowanie się w różnych językach oznacza nieustanną konfrontację językową i konieczność przekładu. Nierzadko zdarza się, iż tłumaczenie jest problematyczne i z trudem znajdujemy odpowiedniki dla poszczególnych słów czy wyrażeń; dotyczy to zwłaszcza jednostek wieloznacznych, a zwłaszcza czasowników. Kłopotliwe staje się odkodowanie znaczenia danej jednostki w języku wyjściowym, pojawiają się też komplikacje w jej realizacji w języku docelowym. Problemy ze zrozumieniem i przekładem czasowników polisemicznych nie dotyczą jedynie języków odległych genetycznie, typologicznie czy geograficznie; trudności mogą występować również w przypadku języków blisko spokrewnionych – np. języka czeskiego i polskiego (Kaczmarska i Rosen 2015b). Badania wykazały, iż często nie jest możliwe odnalezienie jednego jednowyrazowego czasownikowego ekwiwalentu pokrywającego wszystkie znaczenia (Lewandowska-Tomaszczyk 2010, 2013; Kaczmarska, Rosen 2013, 2014a; Kaczmarska 2014a, 2014b, 2019); zdarza się, iż odnajdujemy cały szereg czasowników, różniących się znaczeniowo, ale pretendujących do miana ekwiwalentu danej jednostki¹. Wykorzystując w przekładzie takie bliskie, ale jednak „niekompletne” odpowiedniki, tracimy część treści lub modyfikujemy znaczenie, a tłumaczenie staje się uboższe lub mniej wierne (Kaczmarska 2014b: 26, 2019: 9–11). Niezwykle trudno jest bowiem wyrazić dokładnie tę samą treść w drugim języku z uwzględnieniem rejestru, stylu, frekwencji; czasami jest to zadanie wręcz niemożliwe do wypełnienia. Do jednostek sprawiających szczególne problemy należą czasowniki wyrażające różne stany psychiczne, które w większości są wieloznaczne. W przypadku tych czasowników ryzyko modyfikacji znaczenia przekładu wobec oryginału jest szczególnie wysokie, ponieważ jednostki te bywają problematyczne już na etapie rozumienia (Kaczmarska 2012: 260). Pomocne mogłyby być słowniki, jednakże klasyczne słowniki (drukowane) ze względu na swoją objętość² nie mogą zawrzeć wszystkich znaczeń wyrazów wieloznacznych wraz z kontekstami i przykładami³, a te

¹ Na temat ekwiwalencji czeskich i polskich przymiotników w oparciu o dane korpusowe pisał natomiast Zasina (2016).

² Dwutomowy słownik polsko-czeski (Oliva 1994, 1995) zawiera 80 tysięcy haseł z języka literackiego i potocznego, z bogatą frazeologią z uwzględnieniem terminów specjalistycznych – technicznych, biologicznych, z nauk społecznych. Słownik czesko-polski (Siatkowski, Basaj 2002) zawiera 53 tysięcy haseł i 28 tysięcy wyrażeń i zwrotów współczesnego języka czeskiego.

³ Klasyczne dwujęzyczne papierowe słowniki nie zawierają też kompletnych informacji o walencji czasowników; często też brakuje przy hasłach przykładów użycia danego czasownika

w omawianych przez nas przypadkach byłyby konieczne (Kaczmarska 2014a: 192). Szeregi ekwiwalentów wraz z przykładami i (szerokimi) kontekstami oferuje korpus równoległy; takim jest właśnie InterCorp i na podstawie jego danych przeprowadzimy analizę wybranych czasowników. W tym artykule chcielibyśmy przybliżyć problem przekładu wieloznacznych czeskich czasowników na język polski i zwrócić uwagę na ich nieczasownikowe ekwiwalenty wybierane przez tłumaczy.

Kiedy myślimy o przekładzie czasowników pomiędzy językami blisko spokrewnionymi, oczekujemy pojawienia się w języku docelowym również czasownika. Dane korpusowe ukazują jednak, iż często zamiast oczekiwanego czasownika w języku docelowym – tu w języku polskim – pojawia się inna część mowy lub fraza werbo-nominalna, zmieniająca całkowicie strukturę oryginalnej wypowiedzi. Nasuwa się tym samym pytanie: co sprawia, iż tłumacze wybierają wyrażenie strukturalnie odmienne od jednostki w języku wyjściowym oraz co determinuje wybór tłumacza podczas przekładu jednostki wieloznacznej.

W niniejszym opracowaniu nawiązujemy do wcześniejszych badań nad czeskimi czasownikami, które mają w języku polskim ekwiwalenty czasownikowe⁴. Tym razem spróbujemy odnaleźć reguły pojawiania się różnych – również tych nieczasownikowych – rodzajów ekwiwalentów, korzystając z danych uzyskanych z korpusu równoległego InterCorp.

InterCorp

InterCorp (Čermák i Rosen 2012, Rosen i Vavřín 2014) (Kaczmarska, Rosen 2014b) (Rosen, Vavřín, Zasina 2017) jest częścią Czeskiego Korpusu Narodowego i jako projekt powstał na Uniwersytecie Karola w Pradze w 2005 roku⁵. InterCorp to obszerny, ciągle rozbudowywany, synchroniczny korpus paralelny, zawierający różne teksty (beletrystyka, teksty publicystyczne, administracyjne) w 40 językach (stan na listopad 2018). Ekwiwalentne teksty są posegmentowane i wiązane⁶.

Korpus równoległy jest też cenną wielojęzyczną bazą danych do analiz teoretycznych i leksykograficznych, projektów dotyczących nauki języków obcych

w wszystkich podanych znaczeniach. Np. czasownik *toužit* przekładany jest w słowniku jako *tęsknić*, *marzyć*, *pragnąć* (Siatkowski, Basaj 2002). Informacja dotycząca walencji jest dość ograniczona: *toužit* (*po čem, kom*) – *tęsknić* (*za czym, kim*); *pragnąć* (*czego, kogo*); *marzyć* (*o czym, kim*). Nie pojawia się też żaden przykład; brak też wskazówek co do obiektów łączących się z tymi czasownikami.

⁴ Badania te również oparte były na danych z korpusu równoległego InterCorp; wykorzystaliśmy przy nich metody manualne, automatyczne i stochastyczne.

⁵ Źródło i dostęp www.korpus.cz/intercorp.

⁶ Termin „wiązanie tekstów” za Lewandowska-Tomaszczyk (2005).

i badań translatorskich (Kaczmarska 2014a: 193, 2019: 43). Dzięki swoim narzędziom InterCorp jest też rzetelną podstawą analiz statystycznych⁷.

Korpus InterCorp może również służyć jako źródło ekwiwalentów (Kaczmarska 2019). Konkordancje dwu- lub wielojęzyczne zawierające dane czasownik dostarczają nam informacji dotyczących: ram walencyjnych w analizowanych językach, rodzajów argumentów obecnych w tych ramach oraz odcieni znaczeniowych ekwiwalentów. InterCorp umożliwia również wygenerowanie dwujęzycznego słownika dzięki wiązaniu słowo-do-słowa (*word-to-word alignment*)⁸.

Analiza

Badaniu poddamy czeski czasownik *postrádat*⁹. W języku polskim nie istnieje takie pojęcie: niełatwo je odkodować i trudno je oddać w języku polskim. W poświadczeniach pojawia się szereg jednostek zbliżających się do miana ekwiwalentu – część z nich to pojedyncze czasowniki, część z nich to inne struktury.

Z InterCorpu wyekstraktujemy paralelne czesko-polskie konkordancje zawierające wybrane czeskie czasowniki. Następnie dokonamy manualnej identyfikacji polskich ekwiwalentów – zarówno pod względem ich cech syntaktycznych, jak i semantycznych. Materiał językowy eksцерpujemy z trzonu¹⁰ korpusu InterCorp. Poświadczenia wyszukujemy z tekstów pisanych oryginalnie po czesku i ich polskich tłumaczeń, a także z tekstów pisanych oryginalnie po polsku i ich czeskich tłumaczeń.

Co znaczy *postrádat*?

Słownik języka czeskiego (*Slovník spisovného jazyka českého* – dalej SSJČ) definiuje czasownik *postrádat* jako¹¹

⁷ Na temat możliwości wykorzystania InterCorpu: Čermák, Corness, Klégr 2010, Čermák, Kocék 2010, Kaczmarska, Rosen 2013, Kaczmarska 2015b, 2019, Čermák i Nádvořníková 2015. Por. też: <http://www.korpus.cz/intercorp/?req=doc:uvod> [dostęp: 30.10.2018].

⁸ Na temat ekstrakcji słowników dwujęzycznych z korpusu równoległego: Och i Ney 2003, Skoumalová 2008, Jirásek 2011, Kaczmarska, Rosen 2013. Automatyczne wygenerowanie słownika dwujęzycznego na podstawie danych z korpusu InterCorp jest też możliwy za pomocą narzędzia Treq, dostępnego na stronach Czeskiego Korpusu Narodowego www.treq.korpus.cz: Vavřín, Rosen, 2015, Škrabal, Vavřín, 201.

⁹ Czasownik ten był obiektem naszego zainteresowania również we wcześniejszym badaniu (Kaczmarska, Rosen 2018).

¹⁰ *Trzon* korpusu InterCorp (ang. *core*, czes. *jádro*) stanowią teksty ręcznie wiązane, przeważnie beletrystyczne. Oprócz nich korpus obejmuje także automatycznie opracowane teksty – *kolekcje* (Kaczmarska, Rosen, 2014b: 207–209).

¹¹ <http://ssjc.ujc.cas.cz/search.php?hledej=Hledat&heslo=postr%C3%A1dati&sti=EMPTY&w=hesla&hsubstr=no>

1) *postrádat*₁ (koho, co; przestarz. i książk. – *koho, čeho*) odczuwać nieprzyjemnie stratę, nieobecność, nieistnienie kogoś, czegoś (pożądanego):

Malé děti postrádají těžce matku.

Škoda, že nepřijdete, budeme vás postrádat.

2) *postrádat*₂ (książk. o rzeczy) nie mieć czegoś, co powinno być, jakiś niedostatek jest wadą, defektem, nie mieć czegoś w ogóle, być bez czegoś:

Skladba postrádá původní myšlenku a umělecké hodnoty.

Přednáška postrádala dobrý přednes.

Postrádá vkus, zkušenosti a smysl pro humor.

Nepostrádal odvahu.

3) *postrádat*₃ w połączeniu z czasownikiem *mocť* – *mocť postrádat* (koho, co) – obejść się bez kogoś, czegoś; wyrzec się kogoś, czegoś:

Nemůžeme vás postrádat.

Necháme doma vše, co můžeme postrádat.

*Snad byste tu knihu mohl týden postrádat*¹²

Z tymi definicjami zgodne są polskie odpowiedniki proponowane przez słownik klasyczny (Siatkowski, Basaj 2002: 546): 1) odczuwać brak (kogo, czego), cierpieć z powodu braku (kogo, czego); 2) nie mieć (czego). W słowniku tym, podobnie jak w SSJČ, obecna jest również fraza z czasownikiem *mocť* – *móc obejść się* (bez kogo czego). Wśród tych ekwiwalentów nie ma jednak żadnego, który byłby prostym czasownikiem, jednowyrazowym określeniem.

Charakterystyka walencyjna czasownika *postrádat*

Charakterystykę gramatyczną analizowanej jednostki ustalamy na podstawie słowników walencyjnych języka czeskiego VALLEX¹³.

postrádat^{impf}

① trpět nepřítomností

frame ACT₁^{obl} PAT₄^{obl}

example postrádal svou přítelkyni

recipr ACT-PAT: Petr a Marie se vzájemně velmi postrádají

diat deagent: postrádá se zaběhlý pes

passive: Budeme se věnovat hlavně rozšíření akcí pro děti, ty byly nejvíce postrádány. Josefa se zúčastňovala většiny z nich, a pokud jí v tom okolnosti zabránily, bývala slečna Pepička citelně postrádána.

¹² Definicje i przykłady za SSJČ.

¹³ Dostęp online: <http://ufal.mff.cuni.cz/vallex/3.0/#/lexeme/postr1/0>.

② **nemít**frame **ACT**₁^{obl} **PAT**_{2,4}^{obl}

example skladba postrádá původní myšlenku; postrádal chuť něco dělat; bude postrádat jeho dobrých rad

diat deagent: pozdě večer se již postrádá chuť do práce; postrádá se k tomu chuti

oraz PDT-Vallex¹⁴**postrádat****postrádat**¹_{2x, 17x} **ACT**(1;f) **PAT**(4)(nemít) • *postrádat humor; p. smysl pro legraci***postrádat**²_{31x, 11x} **ACT**(1) **PAT**(4)(nemít) • *postrádat dva hráče*

Czasownik *postrádat* otwiera miejsce dla argumentu o wartości ACT (Aktor), wyrażonego na powierzchni frazą w mianowniku oraz dla argumentu PAT (Patiens, wyrażonego w strukturze powierzchniowej frazą w bierniku (wyjątkowo w dopełniaczu). Ponieważ rama walencyjna czasownika może odgrywać rolę przy wyborze ekwiwalentu (Kaczmarska 2019: 128–133), przyjrzymy się również strukturze walencyjnej polskich odpowiedników wyekscerpowanych z InterCorpu.¹⁵

Automatyczna ekscerpja z korpusu InterCorp i ręczna analiza wyrównanych segmentów

• Czeskie oryginały w tłumaczeniu na język polski

Wyszukując analizowany czasownik w tekstach oryginalnie pisanych w języku czeskim, znaleźliśmy 19 poświadczeń.

CZ oryginały → PL	V	nV
<i>brak</i>		3
<i>brakować</i>	9	
<i>nie mieć</i>	2	
<i>pozbawione</i>		1

¹⁴ Dostęp online: <http://lindat.mff.cuni.cz/services/PDT-Vallex/PDT-Vallex.html>.

¹⁵ Wymagania walencyjne – składniowe i semantyczne zarówno jednostki w języku wyjściowym, jak i potencjalnych odpowiedników w języku docelowym mogą wpływać na decyzję tłumaczy przy wyborze ekwiwalentu (Kaczmarska, Rosen 2013).

<i>utracić</i>	1	
<i>nie zauważyć</i>	1	
POMIENIĘCIE		2
Razem	13	6

Trzyńście z nich to proste czasowniki (jednowyrazowe):

*Nebude vás barman **postrádat**?*

*A barmanowi nie będzie pani **brakowało**?*

*Prokop byl vskutku nadprůměrně inteligentní, i když například **postrádal** nějaký opravdový talent.*

*Prokop rzeczywiście był nieprzeciętnie inteligentny, chociaż na przykład **nie miał** żadnego autentycznego talentu.*

*Lidé říkají, že tisíce rodin, které své mužské příslušníky ve válce buď ztratily nebo je nadlouho **postrádají**, nyní přinášejí své žaloby a nářky na stupně trůnu.*

*Ludzie powiadają, że rodziny, które już utraciły lub lada dzień **utracą** na wojnie swych męskich przedstawicieli, zanoszą teraz skargi i lamenty do stopni tronu.*

Dalsze 4 poświadczenia to nieczasownikowe struktury:

*Zdálo se mu, že **postrádá** zápachu pízma a benátského mýdla.*

*Wydało mu się, że **brak** tu tylko zapachu pizma i weneckiego mydła.*

*Strašlivě jsem **postrádal** nějaký použitelný citát.*

*Odczuwałem dotkliwy **brak** jakiegoś odpowiedniego cytatu.*

*Smyslu může **postrádat** jen chování šilence či výjimečné jednání v prudkém hnutí myšli.*

***Pozbawione** sensu może być tylko zachowanie się wariata albo jakieś dziwaczne działanie podyktowane gwałtownym uczuciem czy niezwykłym poruszeniem myśli.*

Dwa razy pojawiło się też pominięcie w tekście polskim analizowanego czasownika:

*Bez ohledu na to, že v továrně pracovali přes třicet let a mají na rozdíl od mladých kvalifikaci, kterou v továrně **postrádáme**. Bez ohledu na to, že jsou to naši bývalí přátelé ... Bez względu na to, že pracowali w fabryce ponad trzydzieści lat i – w odróżnieniu od młodych – posiadają kwalifikacje. **???** Bez względu na to, że są to nasi dawni przyjaciele ...*

*Osobně bych brala míň bonmotů a víc elementární lidský zdvořilosti jestli víš, co to je. Přesně to u tebe totiž **postrádám**. Slyším dobře?*

*Osobiście wołałabym mniej bon motów, a więcej podstawowej, ludzkiej uprzejmości, jeśli wiesz, co to jest. **???** Czy ja dobrze słyszę ?*

• Polskie oryginały w tłumaczeniu na język czeski

W czeskich tłumaczeniach polskich tekstów znalazły się 34 poświadczenia czasownika *postrádat*.

PL oryginały → CZ	V	nV
<i>bez (przekonania)</i>		1
<i>bez (sensu)</i>		1
<i>bez (znaczenia)</i>		1
<i>bezsensowne</i>		2
<i>brak</i>		2
<i>brakować</i>	2	
<i>nie być</i>	2	
<i>nie dostawać</i>	1	
<i>nie mieć</i>	4	
<i>nie posiadać</i>	1	
<i>nie znajdować</i>	1	
<i>nie(lotny)</i>		1
<i>pozbawiony</i>		9
<i>wątpić</i>	1	
<i>zaginiony</i>		1
INNE		3
NADMIAR		1
Razem	12	22

Dwanaście z nich to przekład polskich czasowników prostych (jednowyrazowych), przy czym wszystkie (poza *brakować*) występują z zaprzeczeniem:

*Ani proti mně nemůžete použít svého oblíbeného argumentu, totiž tvrdit, že **postrádám** cit pro poezii, protože já ho mám, a to v dosti velké míře...*

*A również nie możecie użyć przeciwko mnie waszego ulubionego argumentu, twierdząc że **nie posiadam** wrażliwości poetyckiej, albowiem właśnie posiadam ją i w dużej mierze...*

*[...] sny otce starého K. stále **postrádaly** reálnou předlohu.*

*[...] sny ojca starego K. wciąż **nie znajdowały** swojej jawnej analogii.*

*Jinak by letadlo **postrádalo** smysl.*

*Inaczej ten samolot **nie miałby** sensu.*

*Takovou Annu **postrádal**.*

***Brakowało** mu tamtej Anny.*

Osiemnaście razy pojawił się analizowany czasownik jako przekład różnych polskich struktur nieczasownikowych:

*[...] nalezení filmu z norských fjordů, který dávno **postrádala** a netušila, že se skrýval v nejbzdálenějším koutku nejvrchnější poličky...*

*[...] odnalezienie w najdalszym kącie górnej półeczki dawno **zaginionego** filmu z fjordów norweskich...*

*Nezávislost mají jen žebraři; dokonce i miliardáři ji **postrádají**.*

*Niezależność mają tylko nędzarze, bo już nawet i miliardarzy są jej **pozbawieni**.*

[...] *cítím, že jsou nevhodné, nehezke a neohrabané, a třebaže přinášejí výsledky, naprosto **postrádají** jakýkoli vzlet.*

[...] *są one, w moim odczuciu, nieporęczne, brzydkie, niezgrabne i chociaż przynoszą rezultaty, **czynią** to w sposób całkowicie **nielotny**.*

Trzy poświadczenia dokumentują przekłady, w których czasownik *postrádat* stanowi ekwiwalent kontekstowy i zastępuje inną strukturę:

*Teoreticky jsem tedy zbývala leda já, a aby čistě kvůli mně svítila velká lampa, to mi připadalo vážně jako dost marnotratné plýtvání elektřinou. Proto jsem tam Alici opravdu **postrádala**. Jenže naše hostitelka neustále kroužila mezi kuchyní a terasou a s masochistickou tvrdošijností obsluhovala celou sešlost.*

[...] *pozostawałam ja, ale na mnie jedną marnować całą lampę to doprawdy zbyt duża rozrzutność! **Alicja stanowczo powinna...** Alicja jednak bez chwili przerwy krążyła pomiędzy kuchnią a tarasem, z masochistycznym uporem obsługując towarzystwo.*

[...] *veškeré tyto kombinace jsou odsouzeny k nezdaru, protože **postrádají** trvalost.*

[...] *takie kombinowanie jest klęskowe przez to, że **skazuje na krótkotrwałość**.*

W jednym poświadczeniu czasownik wprowadza frazę, której nie ma w polskim oryginale, jest nadmiarowy:

*Nemusela jsem se ptát, co to je, neboť jsem okamžitě uhodla, co Alice drží v rukou: rodinné dokumenty, které už dlouho **postrádala** a z jejichž zmizení podezírala Anitu. Gratuluju, z obou stránek výborná kombinace, neodpustila jsem si pochvalnou poznámku. Nie musiałam pytać, zgadłam od razu, cholerne dokumenty rodzinne, **???** o które posądzala Anitę! – Obie możliwości idealnie właściwe! – wyrwało mi się pochwalnie.*

Rezultaty

Badanie wykonane na podstawie czeskich oryginałów i przekładów na język polski nie przyniosło żadnych zaskakujących rozstrzygnięć; liczba wyekscerpowanych poświadczeń była zbyt niska. Przykłady ukazały, iż w zgromadzonym materiale tłumacze wybierali najchętniej czasownik *brakować* lub leksem *brak* dla wyrażenia treści zawartej w czasowniku *postrádat*.

Druga część analizy, wykorzystująca polskie oryginały i ich przekład na język czeski, dotyczyła zaledwie 34 poświadczeń. Badanie to jednak zasygnalizowało pewną regułę pojawiania się czasownika *postrádat* jako tłumaczenia. Czasownik ten bowiem często pojawia się w przekładzie tekstu zawierającego w języku polskim prefiks *bez-* lub słowo *bez* (jeśli jednostka w języku polskim jest wielowyrazowa), np. *bezsensowna*, *bez sensu*, *bez przekonania* – jak w przykładach poniżej:

*Na obou stranách se šířil zmatek – kupříkladu osmdesátý sbor marlabardníků prohlásil, že pojem nepřítel není definován dostatečně přesně a obsahuje logické kontradikce, takže **postrádá vůbec smysl**.*

*Po obu stranach robilo się niejaki zamieszanie; osiemdziesiąty korpus marlabardzki wołał, iż trzeba koniecznie uściślić definicję pojęcia „nieprzyjaciel”, która wydaje się, jak dotąd, obciążona logicznymi sprzecznościami, a może nawet zgoła **bezsensowna**. Doktor jí doporučil, aby nedělala žádné prudké pohyby, nyní to však **postrádalo** jakýkoli **mysl**.*

*Lekarz mówił jej, żeby nie robiła gwałtownych ruchów, ale teraz było to **bez znaczenia**. Strašivec to zkouší znovu, tentokrát na jiné frekvenci – zase bez úspěchu, protože už **postrádá** ten pravý **elán**.*

*Spróbowała drugi raz, z innej fazy, ale nie wyszło jej już; **bez przekonania** straszyla.*

Wnioski

Wygenerowane poświadczenia stanowiły zbyt małą ilość materiału, by wyjść w badaniu poza proste porównywanie wyrównanych segmentów i wyszukiwaniu ekwiwalentów. Mając więcej materiału językowego, moglibyśmy użyć metody zaaplikowane do badania czasownika *toužit* – rozszerzone o analizę argumentów badanie walencji jednostek oryginalnych i ich ekwiwalentów, badanie kontekstu i rozróżnienia łączliwości z różnymi obiektami, czy stochastyczne modelowanie wyboru ekwiwalentu na podstawie analizy kontekstu¹⁶ (Kaczmarska, Rosen 2013, Kaczmarska 2015a, 2019; Kaczmarska, Rosen, Hladká 2015). Wykorzystanie szerokiego kontekstu w dużej ilości danych pomogłoby również w identyfikacji strategii translatorskich, włączając w to strategię kompensacji. Wszelkie badania uległyby też znacznemu uproszczeniu, gdybyśmy dysponowali korpusem równoległym z anotacją semantyczną.

Taki semantycznie anotowany, dużo obszerniejszy równoległy korpus czesko-polski jest jednak kwestią przyszłości. Obecnie, w ramach poszerzenia danych korpusowych, możemy korzystać z wyrównanych segmentów pochodzących z tekstów obcojęzycznych tłumaczonych na język czeski i polski, a zawartych w korpusie InterCorp. Dane te wykorzystuje również narzędzie do wyszukiwania ekwiwalentów przekładowych – Treq¹⁷ (Rosen, Vavřín 2015, Škrabal, Vavřín 2017).

¹⁶ Próby stochastycznego modelowania wyboru ekwiwalentu na podstawie analizy kontekstu polegają na wytypowaniu ekwiwalentu na podstawie badania najbliższego otoczenia. Wykorzystuje się przy tym dwie metody: analizę szeregu kilku leksemów przed i za badanym słowem kluczowym oraz analizę leksemów bezpośrednio zależnych od słowa kluczowego (Kaczmarska 2015b: 153).

¹⁷ Dostęp online: <http://treq.korpus.cz/index.php>



Ten właśnie serwis przynosi cały szereg jednostek ekwiwalentnych również dla analizowanego tu czasownika *postrádat*¹⁸:

▲ Frekwence ▼	▲ Procenta ▼	▲ Čeština ▼	▲ Poliština ▼
224	31.5	postrádat	brakować
118	16.6	postrádat	brak
71	10.0	postrádat	tęsknić
53	7.5	postrádat	pozbawić
26	3.7	postrádat	nie
15	2.1	postrádat	zabraknąć
6	0.8	postrádat	zniknąć
6	0.8	postrádat	zaginać
6	0.8	postrádat	mało
4	0.6	postrádat	zgubić
4	0.6	postrádat	potrzebować
4	0.6	postrádat	dysponować
3	0.4	postrádat	pomijać
3	0.4	postrádat	tęsknota
3	0.4	postrádat	zaginiony
3	0.4	postrádat	nieobecność
3	0.4	postrádat	szukać
2	0.3	postrádat	autorytet
2	0.3	postrádat	zapomnieć
2	0.3	postrádat	stęsknić
2	0.3	postrádat	potrzebny
2	0.3	postrádat	nikt
2	0.3	postrádat	charakteryzować

¹⁸ Te same poświadczenia uzyskać można wpisując zapytanie do KonTextu (wyszukiwarka korpusu InterCorp), jednakże Treq oferuje natychmiastowe generowanie listy ekwiwalentów z dostępem do poświadczeń.

2	0.3	postrádat	<u>iść</u>
2	0.3	postrádat	<u>žadén</u>
2	0.3	postrádat	<u>zgináč</u>
2	0.3	postrádat	<u>utrácić</u>
2	0.3	postrádat	<u>obejść</u>
2	0.3	postrádat	<u>nedorzeczny</u>
2	0.3	postrádat	<u>zorientować</u>
2	0.3	postrádat	<u>strácić</u>
2	0.3	postrádat	<u>pusty</u>
2	0.3	postrádat	<u>potrafić</u>
2	0.3	postrádat	<u>nic</u>
2	0.3	postrádat	<u>dać</u>
2	0.3	postrádat	<u>znajdywać</u>
2	0.3	postrádat	<u>posiadać</u>
2	0.3	postrádat	<u>zateřsknić</u>
2	0.3	postrádat	<u>strata</u>
2	0.3	postrádat	<u>bez</u>

Automatyczna ekscerpcja, na której opiera się Treq, nie wychodzi jednak od tekstów oryginalnie pisanych po czesku; za podstawę – jak zostało to już wspomniane – bierze wszystkie teksty w tym języku (również tłumaczenia). Stąd wysokie liczby poświadczeń (Kaczmarska 2018). Korzystając z narzędzia Treq, należy jednak brać pod uwagę, iż stosowana jest tu metoda automatyczna, a wśród proponowanych przekładowych ekwiwalentów znaleźć się mogą sporadycznie również antonimy czy przypadkowe wyrazy (choćby jak powyżej – partykuła przecząca *nie*). Nie jest więc to typowy słownik, zawierający wyłącznie trafione ekwiwalenty, co zostało szerzej opisane we wcześniejszych pracach (Kaczmarska 2016, Kaczmarska, Rosen 2013).

Korzystanie z narzędzia takiego jak Treq jest szczególnie przydatne w sytuacji, kiedy podczas analizy tekstów oryginalnych nie odnajduje się żadne poświadczenie. Wówczas pomocą służą przekłady z innych języków.

Bibliografia

- Čermák František, Corness Patrick, Klégr Aleš (red.), 2010, InterCorp: Exploring a Multilingual Corpus, Praha.
- Čermák František, Kocek Jan (red.), 2010, Mnohojazyčný korpus InterCorp: Možnosti studia, Nakladatelství Lidové noviny, Praha.
- Čermák František, Rosen Alexandr, 2012, The case of InterCorp, a multilingual parallel corpus, „International Journal of Corpus Linguistics”, No 13(3), s. 411–427.

- Čermák, Petr, Nádvořníková Olga et al., 2015, Románské jazyky a čeština ve světle paralelních korpusů, Praha.
- Jirásek Karel, 2011, Využití paralelního korpusu InterCorp k získávání ekvivalentů pro chorvatsko-český slovník. W: Korpusová lingvistika Praha 2011: 1 – InterCorp, red. František Čermák, 45–55, Praha.
- Kaczmarska Elżbieta, 2012, Czeski czasownik „zdát se” w przekładzie na język polski (na podstawie badań z wykorzystaniem czesko-polskiego korpusu równoległego InterCorp), „Studia z Filologii Polskiej i Słowiańskiej”, t. 47, s. 247–261.
- Kaczmarska Elżbieta, 2014a, Czy na pewno się (nie)rozumiemy? O problemach, uproszczeniach, stratach w przekładzie (na podstawie czesko-polskiej części korpusu równoległego InterCorp), [w:] Proměny polonistiky. Tradice a výzvy polonistických studií, red. Michala Benešová, Renata Rusin Dybalska, Lucie Zakopalová, Praha, s. 192–199.
- Kaczmarska Elżbieta, 2014b, Czeskie czasowniki oznaczające stany psychiczne – sposoby ustalania polskich ekwiwalentów na podstawie korpusu równoległego InterCorp, Zbliżenia: językoznawstwo – translatoryka – literaturoznawstwo, Konin 13–14.11. 2013, wykład konferencyjny.
- Kaczmarska Elżbieta, 2015a, W poszukiwaniu znaczenia czasowników wyrażających stany psychiczne. Analiza czeskich czasowników i ich polskich ekwiwalentów – próba implementacji wybranych teorii lingwistycznych (walencja, gramatyka przypadków głębokich, Pattern Grammar, lingwistyka kognitywna), „Prace Filologiczne”, t. LXXVII, s. 131–150.
- Kaczmarska Elżbieta, 2015b, Mít rád czy milovat? O českéj lásce po polsku, [w:] Pojęcia zapisane w języku, red. Krystyna Waszakowa, Marta Falkowska, Warszawa, s. 139–156.
- Kaczmarska Elżbieta, 2016, O dwóch jednostkach leksykalnych będących wykładnikami negatywnych stanów emocjonalnych i ich polskich ekwiwalentach. Analiza na materiale z korpusu równoległego InterCorp, [w:] Polskojęzyczne korpusy równoległe, red. Ewa Gruszczyńska, Agnieszka Leńko-Szymańska, Warszawa, s. 227–248.
- Kaczmarska Elżbieta, 2018, Walencja w przekładzie – stracona czy zachowana? „Prace Filologiczne”, t. LXXII, s. 115–132.
- Kaczmarska Elżbieta, 2019, Metody ustalania ekwiwalentów czasowników wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp, Warszawa.
- Kaczmarska Elżbieta, Rosen Alexandr, 2013, Między znaczeniem leksykalnym a walencją – próba opracowania metody ekstrakcji ekwiwalentów na podstawie korpusu równoległego, „Studia z Filologii Polskiej i Słowiańskiej”, t. 48, s. 103–121.
- Kaczmarska Elżbieta, Rosen Alexandr, 2014a, Czego nie można wyrazić w języku polskim, czyli o leksykalnych w nim brakach, „Polonica”, t. 34, s. 53–66.
- Kaczmarska Elżbieta, Rosen Alexandr, 2014b, Praktyczny przewodnik po korpusie równoległym InterCorp, W: Praktyczny przewodnik po korpusach języków słowiańskich, red. Milena Hebal-Jeziarska, Warszawa, 207–231.
- Kaczmarska Elżbieta, Rosen Alexandr, 2015, Niedosłowności w dialogu czesko-polskim. Język trzeciego tysiąclecia VIII: (Nie)dosłowność, Kraków 13–14.03.2014, wykład konferencyjny.
- Kaczmarska Elżbieta, Rosen Alexandr, 2018, When a verb does not „mean” a verb – a corpus-based analysis of Czech verbs and their non-verbal equivalents in Polish, Olomouc Linguistic Colloquium (OLINCO), Olomouc 7–9.06.2018, wykład konferencyjny.

- Kaczmarska Elżbieta, Rosen Alexandr, Hana Jirka, Hladká Barbora, 2015, Syntactico-semantic analysis of arguments as a method for establishing equivalents of Czech and Polish verbs expressing mental states, „Prace Filologiczne”, t. LXVII, s. 151–174.
- Lewandowska-Tomaszczyk Barbara (red.), 2005, Podstawy językoznawstwa korpusowego, Łódź.
- Lewandowska-Tomaszczyk Barbara, 2010, Nowe spojrzenie na przekład: podobieństwo, granice ekwiwalencji i rekonceptualizacja, „Lingwistyka Stosowana”, t. 3, s. 9–13.
- Lewandowska-Tomaszczyk Barbara, 2013, Komunikacja i konstruowanie znaczeń w przekładzie. Zbliżenia: językoznawstwo – translatoryka – literaturoznawstwo, Konin 13–14.11. 2013, nieopublikowany wykład konferencyjny.
- Och Franz Josef, Ney Hermann, 2003, A Systematic Comparison of Various Statistical Alignment Models, „Computational Linguistics”, No 29(1), s. 19–51.
- Oliva Karel, 1994, Polsko-český slovník. Díl 1, A–Ó (Tom 1), Praha.
- Oliva Karel, 1995, Polsko-český slovník. Díl 2, P–Ž (Tom 2), Praha.
- Rosen Alexandr, Vavřín Martin, 2014, Korpus InterCorp – čeština, verze 7 z 19. 12. 2014. <http://www.korpus.cz>, Ústav Českého národního korpusu FF UK, Praha.
- Rosen Alexandr, Vavřín Martin, 2015, Treq, <http://treq.korpus.cz>, Ústav Českého národního korpusu FF UK, Praha.
- Rosen Alexandr, Vavřín Martin, Zasina Adrian Jan, 2017, Korpus InterCorp, verze 10 z 1. 12. 2017, <http://www.korpus.cz>, Ústav Českého národního korpusu FF UK, Praha.
- Siatkowski Janusz, Basaj Mieczysław, 2002, Słownik czesko-polski, Warszawa.
- Skoumalová Hana, 2008, Extracting dictionaries from parallel corpora, [w:] Proceedings of The Third Baltic Conference on Human Language Technologies, Kaunas, s. 297–301.
- SSJČ – Slovník spisovného jazyka českého, red. Bohuslav Havránek i in., zv. 1–4, Praha 1960–1971; vyd. 2, Praha 1989.
- Škrabal Michal, Vavřín Martin, 2017, Databáze překladových ekvivalentů Treq, „Časopis pro moderní filologii”, č. 99 (2), s. 245–260.
- Vavřín Martin, Rosen Alexandr, 2015, Treq, <http://treq.korpus.cz>, Ústav Českého národního korpusu FF UK, Praha.
- Zasina Adrian Jan, 2016, Překlad a Korpus. Textová ekvivalence vybraných českých adjektiv, „Studia Filologiczne”, t. 5, s. 25–58.

Streszczenie

Niniejszy artykuł omawia problem przekładu wieloznacznych czeskich czasowników na język polski; szczegółowej analizie poddany jest czasownik *postrádat*. Badanie wykonane na podstawie czeskich oryginałów i ich przekładów na język polski ukazuje, iż w zgromadzonym materiale tłumacze najchętniej wybierali czasownik *brakować* lub leksem *brak* dla wyrażenia treści zawartej w analizowanym czasowniku. Druga część analizy, wykorzystująca polskie oryginały i ich przekład na język czeski, skłania ku przypuszczeniu, iż istnieje pewna reguła pojawiania się czasownika *postrádat* jako tłumaczenia. Czasownik ten pojawia się często w przekładzie tekstu zawierającego w języku polskim prefiks *bez-* lub słowo *bez* (jeśli jednostka w języku polskim jest wielowyrazowa), np. *bezsenowna*, *bez sensu*, *bez przekonania*.