

Elżbieta Kaczmarska, Alexandr Rosen<sup>1</sup>

## **Między znaczeniem leksykalnym a walencją – próba opracowania metody ekstrakcji ekwiwalentów na podstawie korpusu równoległego**

### **1. Cel**

Kontakt z językami obcymi wymaga ciągłej pracy przekładowej, która zwłaszcza w przypadku języków blisko spokrewnionych nierzadko bywa w nieoczekiwany sposób utrudniona (choćby za sprawą „fałszywych przyjaciół”). O ile większych kłopotów nie sprawiają rzeczowniki pospolite (tłumaczenie nazw własnych jest sztuką samą w sobie<sup>2</sup>), o tyle często problemy powodować może samo rozumienie, a co za tym idzie – również przekład czasowników – szczególnie opisujących stan psychiczny. Zdarza się, iż proponowanymi

---

<sup>1</sup> Praca nad tym artykułem została częściowo dofinansowana z projektu Český národní korpus (LM2011023) czeskiego Ministerstwa Szkolnictwa, Młodzieży i Kultury Fizycznej, w ramach programu Projekty dużych infrastruktur dla nauki, badań i innowacji.

<sup>2</sup> O przekładzie nazw własnych – Cieślíkova A., *Jak ocalić w tłumaczeniu nazwy własne*, [w:] „Między oryginałem a przekładem” II, red. Filipowicz-Rudek M, Konieczna-Twardzikowa J., Kraków 1996, s. 311–320; Hejwowski K., *Kognitywno-komunikacyjna teoria przekładu*, Warszawa 2009; Lewicki R., *Obcość w odbiorze przekładu*, Lublin 2000; Zakharkevich A., *Ocalone czy utracone nazwy własne w tłumaczeniach baśni Marii Konopnickiej „O krasnoludkach i sierotce Marysi”?*, [w:] „Między oryginałem a przekładem” XV, red. Brzozowski J., Filipowicz-Rudek M., Kraków 2009, s. 303–318.

w słownikach ekwiwalentami bywają jednostki znacznie różniące się znaczeniem<sup>3</sup>. Nie oznacza to jednak automatycznie, iż w języku źródłowym dany leksem jest wieloznaczny (bądź że mamy do czynienia z homonią); może mieć on znaczenie, które jest nieoddawalne w języku docelowym za pomocą jednego ekwiwalentu. W trakcie przekładu, wybierając jakikolwiek „ekwiwalent” podany w słowniku, tłumacz musi rezygnować z niektórych znaczeń<sup>4</sup>. W typowych słownikach dwujęzycznych brakuje również informacji o walencji; dodatkowo na ogół nie prezentują (z uwagi na oczywiste ograniczenia) wszystkich możliwych znaczeń wraz z przykładami, które mogłyby pomóc w wyborze odpowiedniego ekwiwalentu. Artykuł ten nie koncentruje się jednak wokół definicji samego ekwiwalentu<sup>5</sup>. W pracy tej ekwiwalent (dalej także – odpowiednik) to jednostka tekstowa w języku docelowym zastępująca jednostkę tekstową w języku źródłowym z zachowaniem równoważności semantycznej na poziomie języka pomiędzy tekstem wyjściowym a tekstem docelowym<sup>6</sup>.

Celem artykułu jest sprawdzenie, czy (i ewentualnie w jaki sposób) walencja (rozumiana formalnie i semantycznie<sup>7</sup>) wpływa na przekład leksemów (tu – cza-

<sup>3</sup> O problemie tym również w: Kaczmarek E., *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” 47, 2012, s. 247–261.

<sup>4</sup> Może się zdarzyć, iż rzeczywiste znaczenie leksemu leży na granicy czy styku pól znaczeniowych kilku innych leksemów drugiego języka, a wskazanie jednego ekwiwalentu jest niemożliwe.

<sup>5</sup> Szerzej na temat ekwiwalencji m.in. w: Baker M., *In Other Words: A Coursebook on translation*, London 1992; Catford J. C., *A Linguistic Theory of Translation: An Essay In Applied Linguistics*, London 1965; Dąbrowska-Prokop U., *Mała encyklopedia przekładoznawstwa*, Częstochowa 2000; Gentzler E., *Contemporary Translation Theories*, London – New York, 1993; Hejwowski K., *Kognitywno-komunikacyjna...*, op. cit.; Kielar B., *Nauczanie tłumaczenia: rozważania nad ekwiwalencją i ekwiwalentnym tekstem przekładu*, [w:] *Przyczynki do teorii i metodyki kształcenia nauczycieli języków obcych i tłumaczy w perspektywie wspólnej Europy*, red. Grucza F., Warszawa 1993, s. 139–150; Koller W., *The Concept of Equivalence and the Object of Translation Studies*, „Target” 7, 2, 1995, s. 191–222; Nida E. A., *Dynamic Equivalence In Translating*, [w:] *An Encyclopaedia of Translation. Chinese-English / English-Chinese*, red. Chan S.-W., Pollard D. E., Hong Kong 1995, s. 223–230; Pieńkoś J., *Przekład i tłumacz we współczesnym świecie. Aspekty lingwistyczne i pozalingwistyczne*, Warszawa 1993.

<sup>6</sup> Na podstawie definicji: *Translation is the replacement of textual material in one language (SL) by equivalent textual material in another language (TL)* – Catford J. C., *A Linguistic Theory of Translation: An Essay In Applied Linguistics*, London 1965, s. 20; Toury G., *In Search of Theory of Translation*, Tel Aviv 1980, s. 63; cyt za: Dąbrowska-Prokop U., *Mała encyklopedia...*, op. cit., s. 68–75.

<sup>7</sup> Dębski A., *Semantyczna walencja czasownika w aspekcie konfrontatywnym*, „Biuletyn Polskiego Towarzystwa Językoznawczego” XXXIX, 1982, s. 79–90; Rytel D., *Wybrane*

sowników) oraz zaprezentowanie możliwości oddania w języku polskim pojęć w nim nieistniejących (por. przypis 4), a obecnych w języku czeskim – języku blisko spokrewnionym. W efekcie badań autorzy wypracują dane, które mogą być wykorzystane jako uzupełnienie tradycyjnych dwujęzycznych słowników.

Przedmiotem analizy jest czasownik *toužit*, rozumiany w języku polskim jako *tęsknić, pragnąć, marzyć*<sup>8</sup>. W rzeczywistości jednak czasownik ten przekładany jest w różny sposób, co dokumentują przykłady z korpusu równoległego InterCorp<sup>9</sup>.

InterCorp to zapoczątkowany w 2005 roku projekt Uniwersytetu Karola w Pradze, a zarazem część Czeskiego Korpusu Narodowego. Jest to duży synchroniczny korpus paralelny, obejmujący 32 języki (stan na sierpień 2013). InterCorp zawiera różne teksty jednobrzmiące (literatura piękna, publicystyka, teksty administracyjne), których „parą” jest zawsze tekst czeski (bądź jako oryginał, bądź jako tłumaczenie). Język czeski jest dla InterCorpu językiem kluczowym. Aktualnie on-line dostępna jest wersja 5 (zawiera 543 miliony słów obcych i 97 milionów słów czeskich, w tym tekstów beletrystycznych – odpowiednio 92 i 53 miliony słów) oraz wersja 6 (zawiera 867 milionów słów obcych i 100 milionów słów czeskich, w tym tekstów beletrystycznych – odpowiednio 139 i 62 miliony słów). Wersja ta dysponuje wiązaniem segmentów na poziomie zdania, tagowaniem morfologicznym oraz lematyzacją<sup>10</sup>.

Korpus równoległy ma służyć, między innymi, jako źródło danych do badań teoretycznych, analiz leksykograficznych, poszukiwań studenckich, projektów dotyczących nauki języków obcych, a przede wszystkim – badań translatorskich<sup>11</sup>.

Wykorzystywana w tej pracy część InterCorpu to korpus czesko-polski, zawierający 29,6 milionów słów w polskich tekstach, z czego beletrystyka stanowi 8,4 miliony słów<sup>12</sup>. Materiałem poddanym badaniu będą właśnie przykłady z tego korpusu oraz pary ekwiwalentów wygenerowane automatycznie metodą wyrównania segmentów słownych – *word-to-word alignment*;

---

<sup>8</sup> Oliva K., *Polsko-český slovník*, Praha 1994; Siatkowski J., Basaj M., *Słownik czesko-polski*, Warszawa 2002.

<sup>9</sup> Čermák F., Rosen A., *The case of InterCorp, a multilingual parallel corpus*, „International Journal of Corpus Linguistics” 13 (3), 2012, s. 411–427.

<sup>10</sup> Źródło i dostęp [www.korpus.cz/intercorp](http://www.korpus.cz/intercorp)

<sup>11</sup> *Podstawy językoznawstwa korpusowego*, red. Lewandowska-Tomaszczyk B., Łódź 2005, s. 43.

<sup>12</sup> Na potrzeby tego badania wykorzystywana była wersja 5 InterCorpu. Wersja 6 zawiera 47,6 miliona słów w ramach części czesko-polskiej korpusu, z czego beletrystyka stanowi 12,7 miliona słów.

w nomenklaturze polskiej występuje w tym kontekście również termin *wiązanie tekstów*<sup>13</sup>. Metoda ta umożliwi frekwencyjną analizę problemu zasygnalizowanego w temacie w oparciu o wymogi walencyjne i semantyczne wyekscerpowanych leksemów.

## 2. Analiza

**2.1. Pierwszy etap analizy** obejmował ekscerpcję, którą autorzy ograniczyli do oryginalnie czeskich tekstów beletrystycznych z InterCorpu (1,7 mln słów w języku czeskim).

Chcąc przedstawić rozmiar materiału poddanego ekscerpcji, należy dokonać rozróżnienia na oryginały czeskie, polskie i inne (obce), występujące w języku czeskim i polskim. Dane liczbowe wyglądałyby następująco (wszystkie dane liczbowe oprócz frekwencji czasownika podane w milionach słów):

**Tabela 1.** Liczba słów (w milionach) w beletrystycznych tekstach w czesko-polskiej części korpusu InterCorp (wersja 5) oraz frekwencja czasownika *toužit*

	czeskie oryginały	polskie oryginały	obce oryginały	Σ
w języku czeskim	1 709	1 757	4 917	8 384
w języku polskim	1 783	1 698	4 903	8 384
<i>toužit</i>	248	146	418	812

Wyszukiwanym lematem był, jak to zostało już wcześniej wspomniane, czasownik *toužit*. Podczas ekscerpcji manualnej (w tej fazie analizy) autorzy pomijali w wyszukiwaniu derywaty i wyrazy bliskoznaczne. W korpusie

*problemy opisu walencyjnego języka*, „Studia z Filologii Polskiej i Słowiańskiej” XXVI, 1989, s. 237–247; *Walencja czasownika a problemy leksykografii dwujęzycznej*, red. Rytel-Kuc D., Wrocław 1991; Z. Greń, D. Rytel-Kuc, *Wykorzystanie przekładów literackich w pracy nad dwujęzycznym słownikiem walencyjnym*, [w:] *Problemy teoretyczno-metodologiczne badań konfrontatywnych języków słowiańskich*, red. Běličová H., Nieszczimienko G., Rudnik-Karwatowa Z., Warszawa 1991, s. 69–78; Čermáková A., *Valence českých substantiv*, Praha 2009; Urbańczyk-Adach N., *Wariantywność walencji czeskiego czasownika*, Warszawa 2011; Daneš F., Hlavsa Z., *Větné vzorce v češtině*, Praha 1987.

<sup>13</sup> Ibidem.

równoległym czesko-polskim (czeski oryginał, tłumaczenia na język polski, tylko beletrystyka – dalej: BELcz→pl) znaleziono 248 poświadczeń czasownika *toužit*<sup>14</sup>.

Otrzymane wyniki zostały otagowane pod względem walencji czasownika czeskiego, następnie pogrupowane pod względem walencji i łączliwości z określonym obiektem. Autorzy wyróżnili następujące typy łączliwości:

- *toužit po* Oabstr (obiekt abstrakcyjny)
  - (1) *Netoužím po tomhle slizkém bratrství.*  
*Nie tęsknię za takim oślizłym braterstwem.*
- *toužit po* Ohum (obiekt osobowy)
  - (2) *Jsi krásná, nepřestanu po tobě toužit a bát se tvé krásy, miluji tě až do smrti, ale ještě více tě nenávidím.*  
*Jesteś piękna, nigdy nie przestanę cię pragnąć i bać się twojej urody, Kocham cię aż do śmierci, ale bardziej cię nienawidzę.*
- *toužit po / do* OR (obiekt realny)

<sup>14</sup> Uzasadnione wydaje się w tym miejscu pytanie, czy InterCorp jest dostatecznie reprezentatywny do tego typu badania i czy reprezentacja różnych wzorów walencyjnych czasownika *toužit* w korpusie InterCorp jest porównywalna z danymi z jednojęzycznego korpusu języka czeskiego. Autorzy sprawdzili to na podstawie Czeskiego Korpusu Narodowego. W korpusie syn (dane z 19.03.2013) znajduje się 1568 milionów pozycji, przy czym czasownik *toužit* ma w nich 49 196 poświadczeń, co wynosi 31,35 ipm (*items per million*); współczynnik ten nie jest większy od współczynnika dla InterCorpu (BELcz→pl), który wynosi 145,08 ipm. Można więc przyjąć, iż czasownik *toužit* reprezentowany jest w korpusie InterCorp w odpowiednim zakresie. Różnica wynika z tego, iż w korpusie syn jest o wiele mniejszy udział literatury pięknej, a termin *pozycja* obejmuje również znaki interpunkcyjne.

Druga wątpliwość dotyczy tego, jak wygląda rozkład procentowy różnych wzorów walencyjnych czasownika *toužit* w korpusie jednojęzycznym i InterCorp. Autorzy brali pod uwagę wzory: *toužit po* + Locative, *toužit* + infinitiv, *toužit* + *aby* + S (zdanie). Oprócz obiektu „*aby* + zdanie” badano te wzory po obu stronach czasownika *toužit*. W rezultacie, przy konfrontacji wyników z korpusu syn z ekscerpcją ręczną z InterCorpu (BELcz→pl), okazało się, iż w korpusie syn i InterCorpie pojawia się podobna liczba wystąpień czasownika *toužit* z przyimkiem *po* i obiektem (60,05% vs 55,98%) oraz *toužit* z bezokolicznikiem (31,59% vs 30,89%). Różnica występuje jedynie w przypadku wzoru *toužit* + *aby* + S (3,50% vs 8,88%).

Trzeci problem dotyczyć może faktu, że do pewnego stopnia śledzimy uzus jednego autora (ewentualnie tłumacza). W poddawanej analizie części InterCorpu (BELcz→pl), w tekstach Milana Kundery wszystkie formy czasownika *toužit* pojawiają się siedmiokrotnie częściej niż we wszystkich innych tekstach. Kundera, jak wynika z przedstawionego podziału, używa mniej połączeń *toužit* + *po* (52,05% vs 68,63%), natomiast więcej *toužit* + infinitiv (36,30% vs 21,57%) oraz *toužit* + *aby* (10,96% vs. 8,47%). Różnice nie są diametralne, a wysoką częstotliwość tych form można potraktować jako dobrodziejstwo w kontekście badania tego typu.

*po* + OR

- (3) *Když jsem si to uvědomil, přepadlo mne málem zoufalství: cítil jsem se tu jako trosečník a toužil jsem najednou žíznivě po Praze, po své práci, po psacím stolek u ve svém bytě, po knihách.*

*Kiedy to sobie uprzytomniłem, ogarnęła mnie niemalże rozpacz; czułem się tu jak rozbitek i nagle gorąco zatęskniłem za Pragą, za swoją pracą, za biurkiem w swoim mieszkaniu, za książkami.*

*do* + OR

- (4) *to tělo již toužilo do hrobu*  
*to ciało tęskniło już do grobu*

- *toužit* + inf (bezokolicznik)

- (5) *Každý z nás touží překročit erotické konvence, erotická tabu, a vstoupit v omámení do království Zakázaného.*

*Każdy z nas pragnie przekroczyć erotyczne konwencje, tabu, i w upojeniu wstąpić w królestwo Zakazu.*

- *toužit* + S (zdanie podrzędne<sup>15</sup>)

*toužit, aby*

- (6) *Všechno, co se kolem ní dělo, ji obtěžovalo a rušilo a ona toužila, aby se nedělo nic.*  
*Wszystko, co działo się dookoła, zawadzało jej, męczyło, i pragnęła, by nie działo się nic.*

*toužit po tom, aby*

- (7) *Nesmírně toužila po tom, aby se s ní oženil, ale bála se, že kdyby k tomu došlo příliš záhy, cítil by se spoután a ztratila by ho pak tím jistěji.*  
*Pragnęła bardzo, żeby się z nią ożenił, ale obawiała się, że o ile stanie się to za wcześnie, będzie się czuł spętany, a ona straci go tym prędzej i tym pewniej.*

Każda z grup została następnie przefiltrowana pod kątem polskiego ekwiwalentu (i jego walencji). Wyniki zostały przedstawione w tabelkach ukazujących, „które” *toužit* w jaki sposób jest przekładane.

### 2.1.1. Możliwości rozumienia i przekładu frazy *toužit po* Oabstr (94 poświadczenia)

Jest to najliczniejsza grupa i najbarwniejsza pod względem przekładu. W przypadku jednostki *toužit* łączącej się z obiektem abstrakcyjnym obser-

<sup>15</sup> Czasownik *toužit*, jak wynika ze zgromadzonego materiału, łączy się z frazą zdaniową na dwa sposoby – bezpośrednio – *toužit, aby* oraz pośrednio (ze wskaźnikiem zespolenia) – *toužit po tom, aby*.



wujemy skłonność tłumaczy do wyboru ekwiwalentu *pragnąć* (oraz jednostek bliskoznacznych, np. *pożądać*). Przybliżoną liczbę poświadczeń mają tutaj jednak także czasowniki *marzyć* i  *tęsknić* (wraz z formami bliskoznacznymi). Na uwagę zasługuje fakt, iż w tej grupie znalazło się najwięcej jednostek  *tęsknić* łączących się z przyimkiem *do* jako ekwiwalentów czasownika *toužit*. Pojedynczo pojawiają się także przykłady zawierające inne jednostki (m.in. *dążyć do*, *dybać*, *myśleć*, *szukać*, *upragniony*, *zamierzać*, *żądny*).

- (8) *tam již se ukláněł sám papežský nuncius Karel Caraffa, který tak dlouho toužil po pražských vavřínech, ale probošt již mluvil.*  
*Ach, tam kłaniał się sam nuncjusz papieski Karol Caraffa, który tak dawno pragnął wawrzynów praskich, ale prepozyt już mówił.*
- (9) *Ale zatím chce, abych život snášel a po smrti toužil.*  
*A tymczasem chce, bym życie znosił, a śmierci pragnął.*
- (10) *Toužili po hodnosti císařské, opakoval hrabě a apoštolskému misionáři táhly před očima nejpochemurnější obrazy.*  
*Marzyli o majestacie cesarskim – powtórzył hrabia, a misjonarzowi apostołskiemu przesunęły się przed oczami najstraszniejsze obrazy.*
- (11) *a myslím, že není hříchem, toužíme-li v srdci svém po odpočinku a pokoji. i soudž, že nie jest grzechem, ješlí w głębi serca tęsknimy za odpoczynkiem i spokojem.*
- (12) *Toužila po světě, kde lidé mluví jinou řečí než on.*  
*Tęskniła do świata, w którym ludzie mówią innym językiem niż on.*

### 2.1.2. Możliwości rozumienia i przekładu frazy *toužit* po Ohum (37 poświadczeń)

Zgromadzone dane prezentują różne możliwości przekładu na język polski czasownika *toužit*, łączącego się z obiektem osobowym. Tłumacze najczęściej wybierali jako ekwiwalent jednostkę *pragnąć* (również *zapagnąć*) oraz wyrazy bliskoznaczne (*pożądać*, *mieć ochotę*) – w sumie jednostki te to 64,86% w tej grupie. W przekładach pojawiają się również pozostałe, proponowane przez słowniki odpowiedniki, ale były wybierane zdecydowanie rzadziej (*marzyć* – 5,4%,  *tęsknić* – 29,72%). Pojedynczo pojawiają się także przykłady zawierające inne jednostki (m.in. *żyć miłość*).

Na podstawie przeprowadzonego badania można by stwierdzić, że czasownik *toužit* łączący się z obiektem osobowym jest najbliższy znaczeniowo polskiemu czasownikowi *pragnąć*, a także *pożądać*. Zdajemy sobie jednak sprawę, iż dysponując tak nielicznymi danymi, można spreparować całą regułę w oparciu o język danego tłumacza. Przyjrzyjmy się więc przykładom, a w nich kontekstom, w których czasownik *toužit* się pojawia:

## – afekt i kontekst erotyczny:

- (13) *Měl pocit, jako by právě po ní toužil už po mnoho let, jako by ji celou tu dobu hledal po celém světě.  
Pocul, že to jej właśnie pragnął przez całe lata, że to jej szukał po całym świecie.*
- (14) *Kdysi toužil jen po nových ženách.  
Niegdyś pożydał tylko nowych kobiet.*
- (15) *Toužila po něm.  
Pragnęła jego obecności.*

– sytuacje, w których pojawia się *tęsknić* (często części ciała, które „reprezentują” całą osobę):

- (16) *Oči, po kterých touží, jsou oči Tomáše.  
Oczy, do których tęskni, to oczy Tomasza.*
- (17) *Nemohl jsem najednou vůbec pochopit, proč jsem tak nepřičetně toužil po jejím těle;  
Nie mogłem naraz absolutnie pojąć, dlaczego tak niepoczytalnie pragnąłem jej ciała;*
- (18) *Toužila po svém vlastním těle, náhle objeveném, nejbližším i nejcizejším a nejvíc vzrušujícím.  
Tęskniła do własnego ciała, nagle odkrytego, najbliższego i najbardziej obcego, najbardziej podniecającego.*

– *tęsknić* za osobą (istotą), ale również bez kontekstu seksualnego:

- (19) *Tys neznala Boha, Lucie, ale toužila jsi po něm.  
Tyś nie znała Boga, Łucjo, ale tęskniłaś do Niego.*
- (20) *Proč by měl Goethe toužit po Herderovi?  
Dlaczego zatem Goethe miałby tęsknić do pośmiertnej obecności Herdera?*
- (21) *Dlouhé hodiny se jí v duchu zuřivě vysmíval a dlouhé hodiny po ní zoufale toužil.  
Całymi godzinami naśmiewał się z niej w duchu i całymi godzinami rozpaczliwie za nią tęsknił.*

## – bardzo pragnąć / chcieć:

- (22) *Copak tys nikdy netoužil po dítěti?  
Czy nigdy nie pragnąłeś mieć dziecka?*

## – marzyć jako opcja przekładu:

- (23) *Vždycky jsem toužila po člověku, který by byl prostý a přímý.  
Zawsze marzyłam o człowieku, który byłby prosty i bezpośredni.*
- (24) *Není to proto, že by ženy, které se chtějí stát matkami, toužily po světlovlasých dětech s modrými očima, ale protože existuje nedostatek spermatických bank.  
Bez obaw – przyszłe matki nie marzą o błękitnookim i jasnowłosym potomstwie, po prostu w bankach spermy brakuje materiału.*



### 2.1.3. Możliwości rozumienia i przekładu frazy *toužit po OR* (15 poświadczeń)

W przypadku tej grupy trudno mówić o jakiejś tendencji. Wszystkie ekwiwalenty mają poświadczenia jednocyfrowe. Najwięcej przykładów zawiera jednostkę *marzyć*, ale pozostałe czasowniki, proponowane przez słowniki, są również obecne (*pragnąć*, *tęsknić*).

- (25) *Mladý muž touží po vlastním divadle.*  
*Młody mężczyzna marzył o własnym teatrze.*
- (26) *Celý život jsem toužila po skutečném domově,*  
*Całe życie tęskniłam za prawdziwym domem,*
- (27) *Mé patro, vyprahlé po noci zpola probdělé a zpola neklidně prosněné, toužilo*  
*po jejím vřelém a mrazivě vonném doušku.*  
*Moje podniebienie, wyschnięte po nocy na w półprzemarzonej i na w pół prze-*  
*śnionej, pragnęło jej gorącego, orzeźwiająco wonnego łyku.*

### 2.1.4. Możliwości rozumienia i przekładu frazy *toužit + inf* (80 poświadczeń)

W przypadku tej grupy możemy mówić o wpływie walencji na znaczenie i rozumienie tego czasownika. Czasowniki *pragnąć*, *chcieć* oraz konstrukcje formalnie im bliskie (także derywaty) to 90% wszystkich ekwiwalentów w tej grupie.

- (28) *Netouží být po smrti ani s Paulem ani s Brigitou.*  
*Nie pragnęła również spotkać się z Paulem ani z Brigitte.*
- (29) *Každý z nás touží překročit erotické konvence, erotická tabu, a vstoupit*  
*v omámení do království Zakázaného.*  
*Każdy z nas pragnie przekroczyć erotyczne konwencje, tabu, i w upojeniu*  
*wstąpić w królestwo Zakazu.*
- (30) *Začal znovu skicovat do náčrtníku obrazy, které toužil namalovat.*  
*Zaczął znowu szkicować w notatniku obrazy, które pragnął namalować*
- (31) *Já si taky umím představit, že si člověk touží vzít život.*  
*Ja również mogę zrozumieć, że człowiek chce ze sobą skończyć.*
- (32) *Nemohu než to ještě jednou zdůraznit: netoužila vidět pohlaví cizího muže.*  
*Nie mogę nie podkreślić tego jeszcze raz: nie chciała oglądać organów płcio-*  
*wych obcego mężczyzny.*
- (33) *když on toužil mluvit prostě, ona zarputile mlčela*  
*gdy jemu zachciewało się wulgarnych słów, ona uparcie milczała*
- (34) *Teprve během těch tří dnů s ním mohla být tak, jak s ním vždycky toužila být.*  
*Te trzy dni były jedynymi, które udało jej się spędzić przy ojcu w warunkach,*  
*o jakich zawsze marzyła.*

- (35) *Nakonec toužíte poznat někoho obyčejnýho.  
W końcu tęskni się za tym, by poznać kogoś zwyczajnego.*

### 2.1.5. **Możliwości rozumienia i przekładu frazy *toužit, aby / po tom, aby*** (23 poświadczenia)

Dość wyraźne znaczenie wydaje się mieć czasownik *toužit* łączący się ze zdaniem podrzędnym. Przeważająca liczba przykładów zawiera czasownik *pragnąć* (lub *chcieć*), *marzyć* pojawia się dwa razy, natomiast *tęsknić* w zebranym materiale nie pojawia się w ogóle.

- (36) *Toužila tehdy, aby už jednou skončila ta nebezpečná cesta zrad.  
Pragnęła wówczas, żeby raz wreszcie skończyła się ta niebezpieczna droga zdrad.*
- (37) *Toužila, aby přišel, toužila, aby ji k sobě pozval!  
Pragnęła, żeby przyszedł, pragnęła, aby zaprosił ją do siebie!*
- (38) *Nesmírně toužila po tom, aby se s ní oženil.  
Pragnęła bardzo, żeby się z nią ożenił.*
- (39) *Touží, aby ji někdo odnaučil být anachronická!  
Marzy, by ktoś nauczył ją, jak nie być anachroniczną!*
- (40) *Ion toužil po tom, aby mu nejvyšší soudy stranily a aby mu ústavní soudci byli po vůli.  
Jemu również marzył się odgórnie sterowany wymiar sprawiedliwości – na przykład pod postacią uległych mu sędziów Trybunału Konstytucyjnego.*

**2.2. Drugi etap analizy** obejmował ekscerpcję z polskich oryginałów i ich przekładów na język czeski. Liczba słów w polskich tekstach oryginalnych jest w przybliżeniu równa liczbie słów w oryginałach czeskich (1,7 mln słów). Poszukiwane były polskie frazy, których czeskim ekwiwalentem w zebranym materiale był czasownik *toužit* (znaleziono 149 poświadczeń<sup>16</sup>). Dalsza metoda postępowania była identyczna jak w pierwszym etapie. Prowadząc badanie w odwrotnym kierunku (polski oryginał i jego przekład na język czeski), autorzy próbowali wyszukać, jakie jednostki z języka polskiego są przekładane jako czasownik *toužit*.

#### 2.2.1. ***toužit po Oabstr*** (57 poświadczeń)

W ponad połowie przykładów w polskiej wersji (oryginał) pojawia się jednostka *pragnąć* (oraz *chcieć*, *pożądać*, *spragniony*). W oryginale występują też m.in.: *marzyć o*, *tęsknić do / za*, *złakniony*.

- (41) *Lid byl unaven, po zdařilé revoluci toužil po klidu a vydechnutí.  
Ludność była zmęczona, po udanej rewolucji pragnęła spokoju i wytchnienia.*

<sup>16</sup> Z całkowitej liczby 149 przykładów 6 nie podlega analizie. Były to błędy celowe autora oryginału polskiego (5 przykładów) oraz błąd wiązania segmentów tekstu (1 przykład).

- (42) *Neuměl jsem toužit po čistých zvucích.  
Nie potrafiłem pożądać czystych dźwięków.*
- (43) *jini její synové nebudou toužit po takových výpravách  
inni jej synowie nie o wyprawach takich będą marzyć*
- (44) *jsme v podstatě hrozně citliví a toužíme po ušlechtilých citech krásných žen...  
jesteśmy bardzo czuli i tęsknimy do dobrych uczuć pięknych kobiet...*
- (45) *Předstíral jsem lásku jako každý samec, který touží po kopulaci.  
Jak každý złakniony kopulacji samiec pozorowałem miłość.*

### 2.2.2. *toužit po Ohum* (7 poświadczeń)

Fraza ta jest najczęściej przekładem jednostek *pragnąć* i *pożądać*.

- (46) *Tolik jsem snila o této chvíli, tolik měsíců jsem po tobě toužila, tolik let jsem na to čekala, tolik jsem tím trpěla.  
Ja tak marzyłam o tej chwili, tyle miesięcy pragnęłam cię, tyle lat czekałam na to, tyle cierpiałam przez to.*
- (47) *O to se opírá princip trhání nebo pálení fotografií osob, které příliš nenávidíme, nebo naopak, po nichž příliš toužíme, než abychom mohli nosit jejich podobenky u sebe. Na tym polega zapewne darcie lub palenie fotografii osób zbyt znieawidzonych lub, odwrotnie, zbyt beznadziejnie pożądaných, by można było stale mieć ich podobizny przy sobie.*
- (48) *A ona vběhla jako Lasička mezi stromy, ukryla se do jejich stínu a odtamtud po něm začala toužiti a vzdychati.  
A ona między drzewa, co na placu były jak Łasica wbiegła i, w ich cień się schroniwszy, stamtąd tęsknić do niego i wzdychać zaczęła.*

### 2.2.3. *toužit po OR* (6 poświadczeń)

W przypadku tej grupy nie można mówić o żadnej wyraźnej tendencji. W polskim oryginale znalazły się bowiem różne jednostki (występujące w podobnych liczbach): *chcieć*, *łaknąć*, *marzyć*, *pragnąć*, *przepadać*.

- (49) *a všichni toužili po vodce, čaji a cigaretách.  
a wszyscy łaknęli wódki, herbaty i papierosów.*
- (50) *Chtěla si koupit šál, po kterém odedávna toužila.  
Chciała kupić sobie szal, o którym od dawna marzyła.*

### 2.2.4. *toužit + inf* (58 poświadczeń)

Większość wyekscerpowanych przykładów z tej grupy w polskim brzmieniu posiada komponent *pragnąć* albo *chcieć*.

- (51) *Někdy už dokonce toužil prostě zavřít oči a zemřít, a nikdy už nevstát z mrtvých.  
Czasami pragnął już nawet, po prostu, zamknąć oczy i umrzeć, i nigdy nie zmartwychwstać.*

- (52) *Horší a menší národy žijí se svými zvířaty a spolu s nimi touží být spaseny. Gorsze i mniejsze narody żyją ze swoimi zwierzętami i razem z nimi chciałyby zostać zbawione.*
- (53) *Řekl, že to pak bývaly velmi vroucné modlitby a že ho nikterak netoužil usmrtit. Powiedział, że były to bardzo żarliwe modlitwy i nie budziło się w nim wówczas pragnienie uśmiercenia jego istoty.*

### 2.2.5. *toužit, aby / po tom, aby* (11 poświadczeń)

Również w tej grupie w polskich wersjach najczęściej występują jednostki *pragnąć* i *chcieć*. Pojawiają się też pojedyncze przykłady z czasownikami *marzyć* i *tęsknić*.

- (54) *Touží jen po tom, aby bylo všechno svědomitě zapsáno a zaplaceno. Pragnie jedynie, by wszystko zostało skrupulatnie odnotowane i zapłacone.*
- (55) *Celý život toužila po tom, aby mohla studovat historii, teprve v důchodu si však udělala dostatek času. Całe życie chciała studiować historię, ale dopiero na rencie znalazła na to czas.*
- (56) *Oni touží po tom, aby se nám nic nepovedlo a nakonec to dopadlo tak, že nám pošlou pohlednici s pozdravy a přáním všeho nejlepšího. Marzą o tym, żeby nic się nam nie udało, a przynajmniej, żeby na koniec wyszło na to, iż przyszła kartka z pozdrowieniami i życzeniami wszystkiego najlepszego.*
- (57) *Kdopak by netoužil po tom, aby se ho dotkla, aby o něho zavadila Neviditelná Ruka? Któż nie tęskni za tym, żeby być dotkniętym, żeby być muśniętym przez Niewidzialną Rękę?*

Reasumując, należy stwierdzić, iż *toužit* pojawia się najczęściej jako przekład polskich konstrukcji z *chcieć*, *pragnąć*, *pożądać*, *tęsknić*, *marzyć*. Warto jednak zwrócić uwagę na fakt, iż często polskie frazeologizmy oraz złożone składniowo i stylistycznie frazy tłumaczone są na język czeski jako *toužit*.

- (58) *Starannie obmyslałam pytanie, które pozwoliło by wydusić z niego więcej, nie dostarczając zarazem żadnej wiedzy. Zato jsem si bedlivě rozmyslela další otázku, neboť jsem toužila vydolovat z něj co nejvíc informací, ale zároveň sama neprozradila pokud možno vůbec nic.*
- (59) *Z rovnou mně odpověděl ostrožností, že již to pevně v tej potřebie Matki naszej serce pocziwe Syna každého do niej, do niej ptakiem się wyrzywa. Odvětil mi stejně opatrně, že ví, že v tuto těžkou hodinu potřebuje naše Matka každého svého Syna, jehož srdce po ní to už i chce vzlétnouti jak Pták.*
- (60) *Jak często i z dumą mawiał ojciec – miałem, czego dusza zapragnie, i nie brakowało mi ptasiego mleka. Jak často a pyśně říkával otec, měl jsem cokoli, po čem jsem toužil, včetně ptačího mléka.*

Po analizie jakościowej oryginalnie polskich tekstów i ich przekładów na język czeski, autorzy postanowili przeprowadzić prosty test frekwencyjny, badający częstotliwość pojawiania się różnych form czasownika *toužit* w przekładach z różnych języków (oraz w czeskich oryginałach tłumaczonych na te języki). Badaniu zostały poddane teksty z języka polskiego, niemieckiego, hiszpańskiego i rosyjskiego. W przypadku każdego języka obcego (dla języka czeskiego) częstotliwość występowania czasownika *toužit* jest mniejsza w przekładach (z tego danego języka) na język czeski niż w tekstach oryginalnie czeskich, por. Tabela 2.

**Tabela 2.** Różnice w użyciu form czasownika *toužit* w zależności od kierunku przekładu w różnych językach.

język obcy		czeski oryginał	obcy oryginał	czeski oryginał / obcy oryginał
polski	wszystkich słów razem	1 786 732	1 723 278	
	liczba <i>toužit</i>	251	145	
	ipm <i>toužit</i>	140,48	84,14	1,67
niemiecki	wszystkich słów razem	2 264 607	4 129 923	
	liczba <i>toužit</i>	291	284	
	ipm <i>toužit</i>	128,50	68,77	1,87
hiszpański	wszystkich słów razem	572 136	6 177 880	
	liczba <i>toužit</i>	129	445	
	ipm <i>toužit</i>	225,47	72,03	3,13
rosyjski	wszystkich słów razem	466 519	765 088	
	liczba <i>toužit</i>	97	67	
	ipm <i>toužit</i>	207,92	87,57	2,37

### 3. Wyniki ekscerpacji manualnej

Podwójna ekscerpacja pozwoliła autorom na zgromadzenie par ekwiwalentów wybranych przez tłumaczy. Różnych odpowiedników czasownika *toužit* znaleziono ponad 30. Tabela 3. prezentuje najczęściej pojawiające się ekwiwalenty wraz z ich wymaganiami walencyjnymi. W pierwszej kolumnie znalazły się

różne odpowiedniki czasownika *toužit* podzielone według walencji polskiego ekwiwalentu. Pozostałe dwie części tabelki przedstawiają liczbę przykładów, w których czeski czasownik *toužit* z walencją określoną w nagłówku kolumny był przełożony za pomocą określonego polskiego czasownika (wraz z jego walencją). Środkowa część tabeli to wynik badania pierwszego etapu, kiedy analizowane były teksty: oryginał – czeski, przekład – polski. Prawa część tabeli prezentuje wyniki badania drugiego etapu: przekład – czeski, oryginał – polski. Ostatnia kolumna, mająca w nagłówku X, ukazuje liczbę poświadczeń, w których czasownik nie był uzupełniony żadną frazą.

**Tabela 3.** Wyniki ekscerpacji manualnej

<i>toužit</i> (po)	val	Σ	Oa	Oh	OR	inf	S	Σ	Oa	Oh	OR	inf	S	X
<i>chcieć</i>	Oa	0						3	2				1	
<i>chcieć</i>	inf	23	1		1	20	1	9	1			6	2	
<i>chcieć</i>	S	2					2	0						
<i>dążyć do</i>	Oa	2	2					0						
<i>łaknąć</i>	OR	0						2			2			
<i>marzyć o</i>	Oa	24	20			4		6	5			1		
<i>marzyć o</i>	Oh	3	1	2				0						
<i>marzyć o</i>	OR	5			5			1			1			
<i>marzyć</i>	S	3			1		2	0						
<i>mieć ochotę</i>	inf	2		1		1		1				1		
<i>pożądać</i>	Oa	5	5					3	3					
<i>pożądać</i>	Oh	5		5				2		2				
<i>pragnąć</i>	Oa	33	29	1		3		26	25				1	
<i>pragnąć</i>	Oh	12		12				2		2				
<i>pragnąć</i>	OR	3			3			1			1			
<i>pragnąć</i>	Inf	53	4	1	1	44	3	24	1			23		
<i>pragnąć</i>	S	15	1			1	13	3					3	
<i>pragnąć</i>		1		1				4						4

<i>toužit (po)</i>	val	Σ	Oa	Oh	OR	inf	S	Σ	Oa	Oh	OR	inf	S	X
<i>pragnienie</i>	Oa+ing	0						3				3		
<i>(s)próbować</i>	inf	0						2				2		
<i>spragniony</i>	Oa	0						3	3					
<i>tęsknić do</i>	Oa	12	11	1				2	2					
<i>tęsknić do</i>	Oh	5		5				1		1				
<i>tęsknić za</i>	Oa	7	7					1	1					
<i>tęsknić za</i>	Oh	2		2				0						
<i>tęsknić za</i>	OR	2			2			0						
<i>złakniony</i>	Oa	0						2	2					
INNE		27	9	7	2	7	2	35	10	2	2	20		1
Σ		219	81	31	13	73	21	145	45	5	4	36	7	5

Wyniki przedstawione w tabeli 3. ukazują, iż wpływ walencji na przekład czasownika *toužit* widoczny jest najbardziej w przypadku łączliwości analizowanego czasownika z bezokolicznikiem; w tekście polskim najczęściej pojawiają się czasowniki *pragnąć* i *chcieć* łączące się również z bezokolicznikiem. Tylko sporadycznie występują w tym miejscu pozostałe odpowiedniki proponowane w słowniku tradycyjnym.

W przypadku łączliwości z pozostałymi obiektami trudno stwierdzić wpływ walencji na rozumienie i przekład tego czasownika.

#### 4. Automatyczna ekstrakcja ekwiwalentów z czesko-polskiego korpusu równoległego

Mając do dyspozycji korpus równoległy czesko-polski, autorzy podjęli próbę automatycznej ekstrakcji par ekwiwalentów; lista takich par to słownik wygenerowany metodą automatyczną<sup>17</sup>.

<sup>17</sup> Por. Jirásek K., *Využití paralelního korpusu InterCorp k získávání ekvivalentů pro chorvatsko-český slovník*, [w:] *Korpusová lingvistika – InterCorp*, red. Čermák D., Praha 2011, s. 45–55.



Wykorzystanym narzędziem była GIZA++<sup>18</sup> – wiązanie segmentów na poziomie wyrazu (*word-to-word alignment*). GIZA++ jest wolno dostępnym programem, wytworzonym w celu statystycznego przekładu maszynowego, który zawiera moduł wiązania segmentów na poziomie wyrazu. Na wstępie czeskie i polskie teksty były w formie lematyzowanej<sup>19</sup>.

W trakcie ekstrakcji wykorzystane zostały teksty z korpusu InterCorp (wersja 6). Uwzględniona została jedynie beletrystyka (bez rozróżnienia na czeskie, polskie czy obce oryginały) – w tym teksty w języku czeskim: 11,885 milionów słów; teksty w języku polskim 11,860 milionów słów. Wiązanie segmentów (zdań) zostało ograniczone do 1:1, co zwiększa wiarygodność wiązania na poziomie zdania i wyrazu.

Metodą tą wyekstraktowano 8,651 milionów par lematów. Po złączeniu identycznych par (z zachowaniem informacji o ich liczbie) powstało 528 tysięcy haseł dwujęzycznych.

**Tabela 4.** Dwanaście najczęstszych ekwiwalentów czasowników *toužit* i *pragnąć* wyekstraktowanych automatycznie (BELcz→pl) z części korpusu InterCorp (wersja 6)

304	<i>toužit</i>	<i>pragnąć</i>	455	<i>pragnąć</i>	<i>chtít</i>
107	<i>toužit</i>	<i>chcít</i>	304	<i>pragnąć</i>	<i>toužit</i>
82	<i>toužit</i>	<i>těsňit</i>	281	<i>pragnąć</i>	<i>přát</i>
70	<i>toužit</i>	<i>maržit</i>	55	<i>pragnąć</i>	<i>rád</i>
40	<i>toužit</i>	<i>požadát</i>	42	<i>pragnąć</i>	<i>snažit</i>
24	<i>toužit</i>	<i>ochota</i>	18	<i>pragnąć</i>	<i>aby</i>
9	<i>toužit</i>	<i>zapragnąć</i>	17	<i>pragnąć</i>	<i>potřebovat</i>
8	<i>toužit</i>	<i>pragnienie</i>	16	<i>pragnąć</i>	<i>touha</i>
8	<i>toužit</i>	<i>těsňota</i>	13	<i>pragnąć</i>	<i>přání</i>
8	<i>toužit</i>	<i>zależít</i>	11	<i>pragnąć</i>	<i>pokoušet</i>
7	<i>toužit</i>	<i>spragniony</i>	11	<i>pragnąć</i>	<i>prahnout</i>
6	<i>toužit</i>	<i>žycít</i>	11	<i>pragnąć</i>	<i>zatoužit</i>

Wyekstraktowane pary porównano z wynikami ekscerpacji ręcznej.

<sup>18</sup> Por. <http://www.statmt.org/moses/giza/GIZA++.html>; Och F. J., Ney H., *A Systematic Comparison of Various Statistical Alignment Models*, „Computational Linguistics” 29, 1, March 2003, s. 19–51.

<sup>19</sup> Do lematyzacji czeskich i polskich tekstów wykorzystano identyczne narzędzia jak w przypadku korpusu InterCorp. Zob. <http://www.korpus.cz/intercorp/?req=page:info>.

**Tabela 5.** Porównanie ekscerpcji manualnej i automatycznej

pol.	cz.-pol.	pol.-cz.	wszystkie	cz.-pol.	pol.-cz.	wszystkie
chcieć	25	14	109	10,50%	12,73%	15,46%
chęć		1	4		0,91%	0,57%
dążyć	2			0,84%		
łaknąć		3			2,73%	
marzyć	36	9	70	15,13%	8,18%	9,93%
ochota	2	1	24	0,84%	0,91%	3,40%
pożądać	10	5	42	4,20%	4,55%	5,96%
pożądany	1	1	2	0,42%	0,91%	0,28%
pragnąć	118	60	308	49,58%	54,55%	43,69%
pragnienie	3	4	8	1,26%	3,64%	1,13%
(s)próbować		2			1,82%	
spragniony		3	7		2,73%	0,99%
szukać	1		4	0,42%		0,57%
tęsknić	32	5	82	13,45%	4,55%	11,63%
tęsknota			8			1,13%
upragniony			5			0,71%
woleć		1	2		0,91%	0,28%
żądny	1		2	0,42%		0,28%
zależec			8			1,13%
zapagnąć	3	1	9	1,26%	0,91%	1,28%
zatęsknić	1		5	0,42%		0,71%
złakniony	2			0,84%		
życzyć	1		6	0,42%		0,85%
INNE	14	8	22	5,88%	7,27%	3,12%
RAZEM	238	110	705	100,00%	100,00%	100,00%

Tabela 5. porównuje wyniki ekscerpcji ręcznej (cz.-pol., pol.-cz.) i automatycznej (wszystkie). Ekscerpcja automatyczna przeprowadzona została w oparciu o wszystkie teksty znajdujące się w czesko-polskiej części korpusu, ale stale była to tylko beletrystyka. Kolumny prezentują liczbę ekwiwalentów czasownika *toužit* przedstawionych w pierwszej kolumnie tej tabeli. Wyniki

ekscerpcji automatycznej odpowiadają w przybliżeniu rezultatom ekscerpcji manualnej. Oznacza to, iż automatyczna metoda ekscerpcji ekwiwalentów przyniosła spodziewany efekt i należy oczekiwać, iż możliwe jest przeprowadzenie kolejnych badań w oparciu o nią. Mimo iż w obu metodach – manualnej i automatycznej – korzystano z nieidentycznych korpusów; do ekscerpcji automatycznej autorzy wykorzystali wszystkie teksty – oryginały i przekłady również z innych języków, jednak ciągle były to teksty beletrystyczne. Dodanie tekstów obcych było wskazane ze względu na to, iż wyniki metody automatycznej są bardziej wiarygodne przy większej ilości danych.

Metoda ta nie bierze jednak pod uwagę walencji, ale opiera się o kontekst. W dalszej fazie można by zamiast par leksemów ekstraktować pary istotnych kolokacji, natomiast w odleglejszej perspektywie wykorzystać syntaktycznie anotowane teksty (równoległy bank drzew – *parallel treebank*).

Automatycznie wygenerowany słownik można wykorzystać jako uzupełnienie konwencjonalnego dwujęzycznego słownika; dla użytkownika znającego obydwa języki, a poszukującego różnych znaczeń danego leksemu automatycznie wygenerowany słownik może mieć duże znaczenie (choćby w przypadku badań translatorskich czy frekwencyjnych). Może się okazać również niezastąpionym narzędziem w badaniach leksemów pewnego typu np. zdrobnień (w tym przypadku korzystniejsze byłoby wykorzystanie sortowania *a tergo*).

## 5. Podsumowanie

Metoda zastosowana do opracowania danej próbki (tu – czasownika *toužit*) może być użyta do badania wszystkich relewantnych leksemów. Autorzy ustalili, iż wpływ walencji na wybór ekwiwalentu jest zauważalny najwyraźniej w przypadku formy *toužit* + infinitiv. Przy pozostałych schematach decydujące znaczenie może mieć wpływ szerszego kontekstu.

Automatyczną metodę ekstrakcji ekwiwalentów, która nie bierze pod uwagę walencji, autorzy postarają się w przyszłości rozszerzyć o ekstrakcję paralelnych kolokacji, która umożliwi bardziej precyzyjne badanie wpływu innych czynników, poza walencją, na wybór danego ekwiwalentu.

## **Between lexical meaning and valency – towards a method for extracting equivalents based on a parallel corpus**

(Summary)

Some entries in translation dictionaries (even in those for closely related languages) offer multiple target equivalents with disparate meanings even when the meaning of the source word is not considered ambiguous within the source language. The translator's task of choosing the best fitting text unit is made even more difficult by sparse exemplification and missing or sketchy valency information. Using a parallel corpus, we explore to what extent valency (seen in a broad sense both as a syntactic and semantic property) can help to cope with a few difficult cases of Czech psych verbs in relation to Polish and how the lexical caps can be filled. Next, we compare the manually compiled lists of translation equivalents in context, annotated and sorted by their valency properties, with results of an automatic method, using word-to-word alignments in a lemmatised parallel corpus to build a list of equivalent pairs, annotated by their frequency. We conclude that valency is an important (though not omnipotent) discriminant for the choice of the target equivalent and that the list of translation pairs extracted automatically from a parallel corpus can be a useful supplement to a standard translation dictionary.