

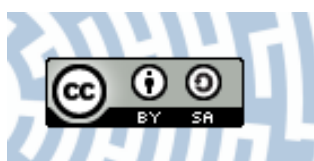


You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice

Title: Treq - narzędzie do ustalania ekwiwalentów przekładowych w korpusie równoległym InterCorp

Author: Andrzej Charciarek

Citation style: Charciarek Andrzej. (2019). Treq - narzędzie do ustalania ekwiwalentów przekładowych w korpusie równoległym InterCorp. W: A. Banaszek-Szapowałowa (red.), "Słowiński krąg : słowo - myśl - obraz w tradycji i współczesności" (S. 293-304). Katowice : Wydawnictwo Uniwersytetu Śląskiego.



Uznanie autorstwa - Na tych samych warunkach - Licencja ta pozwala na kopiowanie, zmienianie, rozprowadzanie, przedstawianie i wykonywanie utworu tak długo, jak tylko na utwory zależne będzie udzielana taka sama licencja.



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

ANDRZEJ CHARCIAREK
Uniwersytet Śląski w Katowicach

Treq – narzędzie do ustalania ekwiwalentów przekładowych w korpusie równoległym InterCorp

Narzędzie Treq¹ jest stosunkowo nowe, pierwsza jego wersja została udostępniona użytkownikom w 2014 r. Jest ono ściśle powiązane z zasobami wielojęzycznego korpusu równoległego InterCorp, stanowiącego część Narodowego Korpusu Języka Czeskiego. Wyniki poszukiwań ewentualnych odpowiedników przekładowych prezentowane są w postaci liczby poświadczeń oraz wartości procentowych. Tak więc tłumacz czy leksykograf już na samym wstępie otrzymuje dane statystyczne, dotyczące użycia danego wyrazu ze wskazaniem potencjalnych ekwiwalentów przekładowych (w kolejności od ekwiwalentu z największą frekwencją do ekwiwalentu z najmniejszą). Każde poświadczenie może być jednocześnie sprawdzone za pomocą odnośnika hipertekstowego, który przenosi użytkownika do konkretnych kontekstów użycia określonego wyrazu. Ponadto wyszukiwanie odpowiedników przekładowych może być zawężone do odpowiednich gatunków tekstów. Zabieg ten umożliwiają zasoby korpusu InterCorp, którego najnowsza wersja 11. z 2018 r. oprócz tekstów beletrystycznych zawiera tzw. kolekcje, czyli teksty publicystyczne i wiadomości ze stron internetowych Project Syndicate i VoxEurop, teksty prawne z korpusu Acquis Communautaire, sprawozdania z obrad Parlamentu Europejskiego (z lat 2007–2011) z korpusu Europarl, napisy filmowe z platformy OpenSubtitles oraz przekłady Biblii (nie są włączone do bazy Treq). Przywołane źródła nie rejestrują tekstów we wszystkich językach, co oznacza, że dla każdej z osobna leksykografii przekładowej, np. czesko-polskiej, czesko-rosyjskiej, InterCorp stwarza odmienne, lepsze lub gorsze, warunki analizy.

¹ Treq jest nazywany również bazą ekwiwalentów przekładowych (czes. databáze překlado vých ekvivalentů, <http://treq.korpus.cz>).

Choć narzędzie Treq czerpie z zasobów korpusu InterCorp, to jednak dla określonego typu użytkownika może ono okazać się nieprzydatne. A to z tej przyczyny, że językami uprzywilejowanymi w tej aplikacji są czeszczyzna i angielszczyzna. Treq kierowany jest głównie do użytkownika czeskiego i anglojęzycznego. Tak więc efektywnie skorzysta z niego czeski rusycysta lub polonista, polski anglista czy bohemista, ale, niestety, dla polskiego rusycysty nie ma on za wiele do zaoferowania – wśród zestawianych języków, tworzących parę przekładową, musi się znaleźć język czeski lub angielski.

Polsko-czeski InterCorp jest jednym z największych (dokładnie szóstym) korpusów w InterCorp, a to za sprawą wielkości modułu polskiego liczącego w ostatniej wersji 11. ponad 86 mln wyrazów. Zawiera on teksty literatury pięknej (ponad 24 mln wyrazów), teksty prawne z korpusu Acquis Communautaire (blisko 20 mln wyrazów), sprawozdania z obrad Parlamentu Europejskiego z korpusu Europarl (prawie 13 mln), teksty publicystyczne i wiadomości ze stron internetowych VoxEurop (prawie 2,5 mln wyrazów), napisy filmowe z bazy OpenSubtitles (około 27 mln wyrazów) oraz przekłady Biblii (ponad 0,5 mln wyrazów). Zwraca uwagę fakt, że w korpusie polsko-czeskim brakuje jednej kolekcji tekstów notowanych w InterCorp, czyli publicystyki ze strony Project Syndicate².

Wyraźnie mniejszy rozmiar ma rosyjski InterCorp v11 (ponad 18 mln wyrazów), który oprócz tradycyjnego rdzenia, czyli tekstów beletrystycznych (ponad 7 mln wyrazów), zawiera teksty publicystyczne ze strony Project Syndicate (niespełna 4 mln wyrazów), napisy filmowe z bazy OpenSubtitles (prawie 7 mln wyrazów) oraz przekłady Biblii (ponad 0,5 mln wyrazów).

Opisane powyżej zasoby polskiego i rosyjskiego korpusu InterCorp nie są, niestety, wykorzystywane przez aplikację Treq, która powiązana jest z mniejszymi zasobami 9. wersji InterCorp z 2016 r.

W przypadku modułu polskiego są one mniejsze o ponad 2 mln wyrazów: 11. wersja – 86 230 000, 9. wersja – 83 822 000. Różnica ta wynika ze znacznego powiększenia się (o prawie 3 mln) liczby tekstów kolekcji literatury pięknej i zmniejszenia się (o ponad 1 mln) kolekcji Acquis Communautaire³.

Rosyjska 9. wersja jest również znacznie mniejsza od 11. wersji. Ta ostatnia, jak już wspomniano, liczy obecnie ponad 18 mln wyrazów, podczas gdy poprzednia wersja 9. – niespełna 15 mln. W wersji 11. przybyło beletrystyki (o ponad 2 mln), tekstów publicystycznych ze strony Project Syndicate (o ponad 0,5 mln wyrazów)⁴.

² Teksty publikowane na stronach Project Syndicate tłumaczone są na język arabski, chiński, czeski, francuski, holenderski, hiszpański, portugalski, rosyjski i szwedzki.

³ Polska wersja 11. notuje ponadto ponad 0,5 mln wspomnianych przekładów Biblii, które nie są przeszukiwane w aplikacji Treq.

⁴ Rosyjska wersja 11. notuje ponadto ponad 0,5 mln wspomnianych przekładów Biblii niedostępnych w aplikacji Treq.

Należy tylko żałować, że Treq nie wykorzystuje pełnego potencjału ilościowego ostatnich polskich i rosyjskich wersji korpusu InterCorp.

Przytoczone dane liczbowe mają istotne znaczenie zwłaszcza w odniesieniu do modułu rosyjskiego, posiadającego w porównaniu z innymi korpusami (nie tylko polskim) niewielkie rozmiary. Wpływa to na nie najwyższą jakość pracy z korpusem czesko-rosyjskim. Przez jakość pracy rozumiemy liczbę pozyskiwanych poświadczeń, która bezpośrednio zależy przecież od wielkości korpusu.

Podsumowując, czesko-polska aplikacja Treq udostępnia użytkownikowi 5 kolekcji, czesko-rosyjska – 3.

W przypadku relatywnie dużych korpusów (ta uwaga dotyczy głównie korpusu czesko-polskiego), użytkownik może poszukiwać ekwiwalentów przekładowych w poszczególnych kolekcjach. Wiadomo przecież, że współczesny język polski, czeski i rosyjski wykazują zróżnicowanie stylowe. Kolekcje rejestrują poszczególne odmiany języka ogólnego, które powiązane są z określonymi sytuacjami komunikacyjnymi. Właśnie w tych odmianach, zwanych inaczej stylami funkcjonalnymi, wykształciły się jedno- lub wielowyrazowe jednostki leksykalne. Są one na tyle charakterystyczne dla poszczególnych stylów, że tworzą odrębne zbiory oznaczane w pracach leksykograficznych odpowiednimi kwalifikatorami. W tym względzie nie ma różnic między polszczyzną, czeszczyzną czy ruszczyzną – każdy z tych trzech języków narodowych jest wewnętrznie zróżnicowany, choć oczywiście, to zróżnicowanie może mieć różny charakter. W każdym z wymienionych języków dostrzeżemy różnice terytorialne, społeczne czy środowiskowe. Zatem możliwość ograniczenia poszukiwań do określonych kolekcji ma istotne znaczenie w kontekście trafności pozyskiwanych ekwiwalentów przekładowych.

Z założeń autorów aplikacji Treq wynika, że wśród ewentualnych ekwiwalentów przekładowych najbardziej adekwatne są te, które posiadają najwyższą frekwencję. Listę ewentualnych translatów uporządkowaną malejąco otwiera jednostka posiadająca najwięcej poświadczeń, zamyka – notowana najrzadziej. Wspomniane ograniczenie przeszukiwanych kolekcji tekstów (funkcja *omezit na*) może skutkować znacznym spadkiem liczby poświadczeń. Fakt ten należy mieć na uwadze, ponieważ mała liczba trafień obniża wiarygodność uzyskanych rezultatów, a niekiedy wręcz udaremnia przeprowadzenie analizy korpusowej.

Wersja 2. narzędzia Treq umożliwia użytkownikom wyszukiwanie nie tylko jednowyrazowców, ale także jednostek wielowyrazowych. To samo dotyczy pozyskiwanych poświadczeń, mogących mieć formę zarówno jednego wyrazu, jak i jednostki wielowyrazowej. Wystarczy w odpowiednim polu (*dotaz*) wpisać jednostkę, dla której poszukuje się ekwiwalentu. Nie sposób nie zwrócić uwagi także na funkcje dodatkowe. Można użyć funkcji *lemmaty* (czes. *lemmata*), co pozwoli na pozyskanie większej liczby ekwiwalentów przekładowych (poprzez wyszukanie wszystkich form fleksyjnych konkretnie-

go leksemu), por. (w kolejności: liczba poświadczeń, udział procentowy, polski transland, czeski translát)⁵:

bez zaznaczania funkcji *lemmaty*

460	57.9	kot	<u>kočka</u>
207	26.1	kot	<u>kocour</u>
39	4.9	kot	<u>kot</u>
15	1.9	kot	<u>kočku</u>
12	1.5	kot	<u>Kočka</u>
6	0.8	kot	<u>Kňour</u>
5	0.6	kot	<u>kočky</u>
5	0.6	kot	<u>Kocour</u>
5	0.6	kot	<u>kocoura</u>
4	0.5	kot	<u>kočce</u>
4	0.5	kot	<u>kočkou</u>
3	0.4	kot	<u>kocka</u>
3	0.4	kot	<u>zvíře</u>
2	0.3	kot	<u>lemoun</u>
2	0.3	kot	<u>Jellikočka</u>
2	0.3	kot	<u>Rufus</u>

z zaznaczeniem funkcji *lemmaty*

1518	67.4	kot	<u>kočka</u>
429	19.0	kot	<u>kocour</u>
47	2.1	kot	<u>kočičí</u>
39	1.7	kot	<u>kóta</u>
17	0.8	kot	<u>koťě</u>
16	0.7	kot	<u>kočička</u>
13	0.6	kot	<u>Kot</u>
12	0.5	kot	<u>kňour</u>
10	0.4	kot	<u>šelma</u>
6	0.3	kot	<u>Kocour</u>
6	0.3	kot	<u>Koček</u>
6	0.3	kot	<u>zvíře</u>
5	0.2	kot	<u>zaječí</u>
5	0.2	kot	<u>lemoun</u>
5	0.2	kot	<u>kočičky</u>

Choć uzyskane wyniki w dużym stopniu wykazują podobieństwa, to nie można nie dostrzec różnic w pozyskanych poświadczeniach.

⁵ Ze względu na brak miejsca nie zawsze prezentuję listy ze wszystkimi poświadczeniami, których może być niekiedy kilkadziesiąt lub kilkaset.

W poszukiwaniu odpowiedników wielowyrzawowych wydatnie pomaga funkcja *wielowyrzawowe* (*viceslovné*). Przy wyborze tej funkcji należy pamiętać o większej liczbie poświadczeń błędnych, o czym zresztą informują sami autorzy aplikacji. Pozyskiwany w ten sposób materiał językowy jest wielce niedoskonały, ale umożliwia – co trzeba mocno podkreślić – wskazanie ewentualnych ekwiwalentów. Wymaga to jednak wnikliwej analizy poszczególnych poświadczeń, przede wszystkim kontekstów ich użycia poprzez wykorzystanie odnośnika hipertekstowego.

Pomocną jest funkcja *regulární*, umożliwiająca bardziej zaawansowane formułowanie zapytań. Tzw. wyrazy regularne to sekwencje symboli, za których pośrednictwem można wyszukać określony zbiór wyrazów. Wyrazy regularne mogą składać się wyłącznie ze znaków specjalnych, być kombinacją znaków specjalnych i alfanumerycznych lub zawierać tylko znaki alfanumeryczne. Funkcja ta jest bardzo przydatna w wyszukiwaniu jednostek wielowyrzawowych. Pozwala mianowicie uwzględnić ważne zjawisko w badaniach korpusowych – wariantywność związków frazeologicznych. Właśnie wychwycenie synonimów komponentów jednostek wielowyrzawowych nastrocza niemało problemów. Użycie różnych wyrazów regularnych może skutecznie pomagać w ich odnotowywaniu.

Najbardziej uniwersalnym znakiem jest kropka (.), która może zastąpić dowolną literę. Tak więc sekwencja trzech kropek z rzędu umożliwi znalezienie wyrazów trzyliterowych, czterech kropek – czteroliterowych itd.

Nie mniej istotnym znakiem jest gwiazdka (*), która zastępuje dowolny ciąg znaków (zero i więcej). Znak plus (+) pełni podobną funkcję, różniąc się od gwiazdki tym, że zastępuje minimum jeden znak lub więcej (jeden i więcej). Z powyższego wynika, iż zastosowanie jednego czy drugiego znaku może przynieść różną liczbę poświadczeń. Np. zadając zapytania *kobiet.+* i *kobiet.**, otrzymamy odmienne wyniki (odpowiednio: liczba poświadczeń, udział procentowy, polski transland, czeski translat):

kobiet.+		kobiet.*					
4677	26.9	kobiety	<u>ženy</u>	4923	19.3	kobiet	<u>žen</u>
3482	20.0	kobieta	<u>žena</u>	4677	18.4	kobiety	<u>ženy</u>
1131	6.5	kobietę	<u>ženu</u>	3482	13.7	kobieta	<u>žena</u>
911	5.2	kobietą	<u>žena</u>	1360	5.3	kobiet	<u>ženy</u>
616	3.5	kobietą	<u>ženou</u>	1131	4.4	kobietę	<u>ženu</u>
517	3.0	kobieta	<u>ženská</u>	911	3.6	kobietą	<u>žena</u>
476	2.7	kobietami	<u>ženami</u>	616	2.4	kobietą	<u>ženou</u>
443	2.5	kobiety	<u>ženu</u>	605	2.4	kobiet	<u>ženách</u>
402	2.3	kobietom	<u>ženám</u>	517	2.0	kobieta	<u>ženská</u>
272	1.6	kobiety	<u>žen</u>	476	1.9	kobietami	<u>ženami</u>

Powyższe zestawienie pokazuje, jak ważne jest prawidłowe zbudowanie zapytania. Znaki o podobnych (ale nie identycznych!) funkcjach odnajdują różną liczbę poświadczeń.

Pytajnik (?) to kwalifikator, który może okazać się pomocny w wyszukiwaniu odpowiedników przekładowych. Zastępuje on jedną literę lub żadną.

Mogą być również użyte inne znaki, pozwalające na wyodrębnienie wariantów jednostek wielowyrazowych: nawiasy okrągłe, nawiasy kwadratowe, odwrócony ukośnik, pionowa linia, daszek. Wszystkie one mogą być użyte w sformułowaniu zapytania.

Czwarta funkcja (A=a) pozwala na nierozróżnianie dużych i małych liter, a więc jej zaznaczenie umożliwi znalezienie np. wyrazów Bieg, bieg, BIEG.

Wydaje się, że narzędzie Treq najlepiej sprawdza się w wyszukiwaniu pojedynczych wyrazów. Proces ten przebiega szybko i sprawnie, w praktyce nie wymaga od użytkownika żadnego specjalistycznego przygotowania, np. znajomości wyrazów regularnych. Uzyskane ekwiwalenty przekładowe z reguły są trafne, a ponadto są (zwykle jest ich więcej niż jeden) automatycznie sklasyfikowane zgodnie z kryterium frekwencyjnym.

Niemniej jednak Treq może posłużyć użytkownikom do zadań bardziej skomplikowanych, w dużym stopniu podobnych do pracy z podstawowym narzędziem korpusu, jakim jest KonText. Uzyskiwane wyniki są jednak prezentowane inaczej, z pewnością mniej przejrzyste. Posłużmy się praktycznymi przykładami, koncentrując się na wyszukiwaniu ekwiwalentów wybranych trzech czeskich frazemów z komponentem *krk* w języku polskim i rosyjskim:

lézt někomu krkem,
mít něco/někoho z krku,
jít někomu po krku.

Dla zapytania *lez.* krkem* wyniki polskich ekwiwalentów wyglądają następująco:

1 6.3	leżla trochu krkem	<u>lubił</u>
1 6.3	leżla krkem	<u>nerwach</u>
1 6.3	leze krkem	<u>mdłości</u>
1 6.3	leze krkem	<u>cierpi</u>
1 6.3	leze krkem	<u>zniosę tej podróbki Elvisa</u>
1 6.3	leze krkem	<u>podróbki</u>
1 6.3	leze krkem	<u>niedobrze tej piosenki</u>
1 6.3	leze krkem	<u>doprowadza</u>
1 6.3	leżla trochu krkem	<u>jej bardzo nie lubił</u>
1 6.3	leżla krkem	<u>na nerwach</u>
1 6.3	leze krkem	<u>gardłem</u>

1 6.3 lezou krkem	<u>wkurzają</u>
1 6.3 leze krkem	<u>słuchać</u>
1 6.3 leze krkem	<u>piosenki</u>
1 6.3 leze krkem	<u>już tego słuchać</u>
1 6.3 leze krkem	<u>bielizny</u>

Jak widać, pozyskane dane korpusowe nie pozwalają na wskazanie ekwiwalentu przekładowego w języku polskim. Treq wskazuje 16 ewentualnych translatów, każdy z nich ma jedno poświadczenie. Jeśli zestawimy je z ewentualnymi odpowiednikami w pracach leksykograficznych (*wyłazić/wychodzić komuś bokiem, mieć kogoś/czegoś po dziurki w nosie*)⁶, to okaże się, że choć ich bezpośrednio nie wskazują, to w miarę skutecznie naprowadzają na dobre rozwiązania ekwiwalentyzacyjne. To, co jest konieczne na pierwszy rzut oka, to rozszerzenie kontekstów użycia, ponieważ Treq podaje zaledwie część (niekiedy jeden wyraz) ewentualnego translatu. Po dokonaniu tej czynności wskazanie potencjalnych translatów jest już o wiele prostsze: *grać komuś na nerwach, doprowadzać kogoś do szału, wkurzać kogoś, wyłazić komuś gardłem, nie znosić czegoś, niedobrze się komuś robić od czegoś, nie móc czegoś słuchać*. Nie wszystkie z nich sprawdzą się jako polskie odpowiedniki przekładowe, choć należy przyznać, że wiele z nich na to miano zasługuje. Co znamienne, niektóre z nich nie są notowane w słownikach przekładowych.

Przejdźmy do czesko-rosyjskiej aplikacji Treq. Oto wyniki wyszukiwań:

1 50.0 lezlo krkem	<u>устал от всего этого дерьма</u>
1 50.0 lezlo krkem	<u>дерьма</u>

Jak można zauważyć, liczba poświadczeń (2) jest więcej niż skromna. Przyczyna tkwi w niewielkiej liczbie (3) i niewielkich rozmiarach przeszukiwanych kolekcji. Po dokładnym sprawdzeniu okazuje się, że Treq w praktyce rejestruje jedno poświadczenie, mylnie rozpoznając je jako dwa odrębne.

Zmiana zapytania na *l.* krkem*, czyli uwzględnienie formy bezokolicznika, przynosi następujące rezultaty:

1 25.0 lézt krkem	<u>тошнит</u>
1 25.0 lezlo krkem	<u>устал от всего этого дерьма</u>
1 25.0 lézt krkem	<u>Меня уже тошнит от этого места</u>
1 25.0 lezlo krkem	<u>дерьма</u>

Po niezbędnej weryfikacji liczba poświadczeń wynosi 2, znów mamy do czynienia z błędnym podwojeniem wyników. Treq wskazuje *тошнить кого-л. от чего-л.* jako ekwiwalent czeskiego frazemu *lézt krkem*. Nawet uznając ten

⁶ *Wielki czesko-polski słownik frazeologiczny*, Kraków 2009, s. 213.

ekwiwalent za trafny, nie można go uznać za reprezentatywny (znikoma liczba poświadczeń) i zaprzestać poszukiwania innych. A te rodzimemu użytkownikowi ruszczyzny zaraz przychodzą na myśl, jak np. *сыт по горло*. W tym miejscu ujawnia się zasadnicza słabość korpusu czesko-rosyjskiego, notującego zbyt małą liczbę tekstów. Należy odnotować brak choćby jednego poświadczenia czeskiego *leze krkem*, które można uznać za kanoniczne.

Frazem *mít něco/někoho z krku* używany jest w znaczeniu ‘nie musieć się zajmować, troszczyć’ i posiada w polszczyźnie odpowiednik słownikowy *mieć kogoś/coś z głowy*⁷. Wyszukiwanie odpowiedników przekładowych przynosi 11 poświadczeń:

1 9.1	mši svatou z krku	<u>mszy</u>
1 9.1	mít tvou přepravu z krku	<u>zmroku załatwić ci przeprawę</u>
1 9.1	máme z krku	<u>najcięższe za</u>
1 9.1	máme z krku	<u>najcięższe mamy</u>
1 9.1	máme z krku	<u>Załatwisz</u>
1 9.1	máme z krku	<u>dla maluchów</u>
1 9.1	mši svatou z krku	<u>dzisiejszej mszy</u>
1 9.1	mít tvou přepravu z krku	<u>przeprawę</u>
1 9.1	máme z krku	<u>najcięższe sobą</u>
1 9.1	mít z krku	<u>wyprawić</u>
1 9.1	máme z krku	<u>maluchów</u>

Podobnie jak w przypadku wyszukiwania translatów *lézt někomu krkem*, pozyskany materiał korpusowy może wprowadzić użytkownika w błąd. Liczba 11 nie odzwierciedla rzeczywistego stanu rzeczy. Po zweryfikowaniu źródeł okazuje się, że w istocie Treq wskazuje 6 poświadczeń. Mamy więc po raz kolejny do czynienia z podwojonymi (utwory Jaroslava Haška i Andrzeja Sapkowskiego), a nawet potrojonymi (powieść *W matni* Emila Zoli) tymi samymi poświadczeniami. Niemniej należy odnotować ewentualne ekwiwalenty przekładowe, które Treq sugeruje: *mieć coś za sobą, załatwić coś*. Owe ekwiwalenty mogą okazać się przydatne w określonych kontekstach. Jednak minusów operacji wyszukiwania translatów za pomocą aplikacji Treq jest więcej. Po pierwsze, ewentualne translaty można wyodrębnić dopiero po rozszerzeniu kontekstu, czyli przejściu do aplikacji KonText, która prezentuje paralelne konkordancje. Po drugie, Treq nie przytacza translatu *mieć kogoś/coś z głowy*, który pod względem stopnia ekwiwalencji jest najbliższy czeskiemu *mít něco/někoho z krku*. Po trzecie, co wynika z poprzedniego, frekwencja pozyskanych poświadczeń jest zbyt mała, aby można było mówić o wiarygodności analizy.

Jeśli zasoby czesko-polskiego narzędzia Treq są większe od czesko-rosyjskiego, to trudno spodziewać się, aby liczba poświadczeń czeskich i rosyjskich

⁷ Ibidem, s. 213.

była większa. Praktyka to potwierdza, czesko-rosyjski Treq notuje 3 poświadczenia:

1 33.3 máme z krku сделаем

1 33.3 máme z krku представится

1 33.3 máme z krku представится шанс позаботиться об этом

W rzeczywistości mamy 2 poświadczenia (jedno jest powtórzone), które nie pozwalają na określenie ekwiwalentu przekładowego. Zresztą wskazanie rosyjskiego odpowiednika *mít něco/někoho z krku* jest trudniejsze, niż określenie jego translatu w polszczyźnie. Nie znajdziemy bowiem we współczesnym języku rosyjskim frazemu z analogicznym komponentem somatycznym. Wydaje się, że ewentualne translaty rosyjskie można wskazać dopiero w zasobach InterCorp za pośrednictwem aplikacji KonText. Jednym z nich, zasługującym być może na pierwsze miejsce po stronie ekwiwalentów, jest z pewnością frazem *с плеч долой*. Znamienne jest, że jest on praktycznie nienotowany w słownikach przekładowych zarówno ogólnych, jak i frazeologicznych⁸.

Trzecim i ostatnim z badanych frazemów jest czeski *jít někomu po krku*. *Wielki czesko-polski słownik frazeologiczny* określa jego znaczenie jako ‘szkodzić komuś’, podając jednocześnie jego dwa polskie ekwiwalenty – *kopać pod kimś dolki* i *podkładać komuś świnie*⁹.

Sekwencja *j.* po krku* posiada 10 poświadczeń, po których zweryfikowaniu pozostaje ostatecznie 5:

1 10.0 jdou vzájemně po krku	<u>gardeł</u>
1 10.0 jde Frankovi kravatou rovnou po krku	<u>szarżuje na Franka</u>
1 10.0 jde Frankovi kravatou rovnou po krku	<u>i szarżuje</u>
1 10.0 jde po krku	<u>prześladuje</u>
1 10.0 jde po krku	<u>dobrał do</u>
1 10.0 jdou vzájemně po krku	<u>do gardeł</u>
1 10.0 jde Frankovi kravatou rovnou po krku	<u>szarżuje</u>
1 10.0 jdou po krku	<u>skaczą</u>
1 10.0 jde po krku	<u>dobrał się</u>
1 10.0 jde po krku	<u>dobrał</u>

Nietrudno zauważyć, że powyższy materiał korpusowy uzyskany za pomocą aplikacji Treq wymaga dalszego opracowania, czyli wejścia do zasobów korpusu InterCorp. Po dokładnej analizie kontekstów tych w praktyce fragmentów frazemów ewentualnymi ekwiwalentami przekładowymi mogą być: *dobrać*

⁸ W *Wielkim słowniku czesko-rosyjskim* znajdujemy dwa ekwiwalenty frazemu przytoczonego w formie *mám to z krku*: это я уже спихнул і с этим я уже развязался (*Velký česko-ruský slovník LEDA*, 2005, s. 331).

⁹ *Wielki czesko-polski słownik frazeologiczny*. Kraków 2009, s. 213.

się komuś do skóry, prześladować, rzucić się do gardła, szarżować. Treq, co ważne, przy tak zbudowanym pytaniu pominie formy supletywne, czyli formy odmiany czasownika *jít* typu: šel jsem, šla jsem, šli jsme itp. Zapytanie š.* *po krku* przynosi kolejne 2 propozycje ekwiwalentów: kontekstowo akceptowalne *šciгаć* i nietrafne *rozwalić gardło*. Poświadczeń jest, jak nietrudno zauważyć, niewiele, ale nawet na ich podstawie można byłoby zakwestionować przywołane wyżej słownikowe pary przekładowe *jít někomu po krku* – *kopać pod kimś dolki*¹⁰ i *jít někomu po krku* – *podkładać komuś świnie*. Na podstawie analizy, poszerzonej z konieczności o zasoby korpusu InterCorp, można stwierdzić, że na miano ekwiwalentów przekładowych bardziej zasługiwałyby: *dobrać się komuś do skóry, rzucić się/skoczyć komuś do gardła, dobrać się do kogoś, dopaść kogoś*. Sytuacja ta (zresztą nie tylko ta) pokazuje, że Treq pomaga w ustaleniu ekwiwalentów, ale częstokroć nie rozstrzyga jednoznacznie wątpliwości, które temu procesowi towarzyszą.

Jeśli liczba pozyskanych poświadczeń polskich odpowiedników *jít někomu po krku* z pomocą Treq jest znikoma, to praktyczny brak ekwiwalentów rosyjskich nie jest żadnym zaskoczeniem. Aplikacja notuje *de facto* 2 (choć wskazuje 4) poświadczenia:

Nebylo to jednoduché a zeshora mi šli po krku. – Но совет директоров дышит мне в затылок.

Už žádný Markus Kane jdoucí ti po krku. – Маркус Кейн больше не станет пытаться убрать тебя.

Powyższy materiał może być co najwyżej punktem wyjścia w procesie ustalenia translatu czy translatów w języku rosyjskim. Nie przesadzając o nieadekwatności wskazywanych przez Treq translatów, znaczenie rosyjskich ekwiwalentów przekładowych *хватать/схватить за горло* i *сесть/садиться на хвост* wykazuje większe podobieństwo do czeskiego translandu. Ale tych translatów narzędzie Treq nie wyszuka, znajdziemy je dopiero po żmudnych przeszukiwaniach zasobów korpusu InterCorp za pomocą aplikacji KonText, a następnie analizie poświadczeń.

Podkreślając niewątpliwą użyteczność aplikacji Treq w procesie przekładu, można sformułować kilka wniosków.

Aplikacja Treq sprawdza się najlepiej w wyszukiwaniu ekwiwalentów przekładowych jednowyrazowców. Dane frekwencyjne, które tą drogą użytkownik uzyskuje, zwykle bezbłędnie wskazują najczęstszy translát. Treq wskazuje też w postaci listy frekwencyjnej inne rzadziej używane transláty, z funkcjonalnymi włącznie. Duża liczba poświadczeń gwarantuje dodatkowo adekwatność translatów.

¹⁰ Wydaje się, że jednym z czeskich ekwiwalentów przekładowych frazemu *kopać pod kimś dolki* jest *strojit úklady*.

Narzędzie Treq może wydatnie pomóc w wyszukiwaniu translatów wielowyrazowców, choć proces ten obarczony jest sporym marginesem błędu. Po pierwsze, Treq wymaga od użytkownika znajomości stosowania określonych symboli w celu prawidłowego zbudowania zapytania. Po drugie, sposób prezentacji znalezionych odpowiedników z reguły zmusza do podglądu źródła, co w praktyce oznacza przejście i korzystanie z aplikacji KonText. Po trzecie, na ogół nieduża liczba poświadczeń (niektóre poświadczenia korpusowe Treq pomija, wiele podwaja, potraja itp.) nie daje pewności, czy wskazywane odpowiedniki są adekwatne. Sytuacje takie praktycznie uniemożliwiają niekiedy przeprowadzenie badań nad określonym zjawiskiem leksykalnym. Po czwarte, wiele odpowiedników wielowyrazowych wskazywanych przez Treq to propozycje błędne.

Nie należy zapominać o tym, że Treq stworzony został do badania języka czeskiego i bazuje na tekstach, które w większości stanowią przekłady, a to oznacza, że najczęściej zestawiane z sobą teksty (w naszym przypadku czeski – polski i czeski – rosyjski) nie są tekstami źródłowymi i tekstami przekładu będącymi efektem tłumaczenia bezpośredniego. Ustanawianie par przekładowych, bo o nie tu chodzi, odbywa się częstokroć w oparciu o tekst w języku trzecim (najczęściej angielskim).

Nie podając w wątpliwość przydatności aplikacji Treq, można tylko żywić nadzieję, że wraz z jego doskonaleniem jako narzędzia wyszukiwawczego i jednoczesnym powiększaniem się zasobów tekstowych korpusu InterCorp stworzy on w przyszłości znacznie lepsze warunki do pracy translatorskiej.

Słowniki

Wielki czesko-polski słownik frazeologiczny. Red. T.Z. ORŁOŚ, Kraków 2009.

Velký česko-ruský slovník LEDA, 2005.

Анджей Харцярек

Treq – программа для установления переводных эквивалентов в параллельном корпусе InterCorp

Настоящая статья посвящена описанию использования корпусной программы Treq в процессе перевода.

Охарактеризованы польский и русский модули параллельного корпуса InterCorp.

Подробно описаны функции приложения Treq.

На конкретных примерах перевода одно- и многолексемных единиц продемонстрированы как преимущества, так и недостатки использования программы Treq.

Ключевые слова: Treq, InterCorp, переводные эквиваленты, чешский язык, польский язык, русский язык

Andrzej Charciarek

**Treq – a tool for searching translation equivalents
in the InterCorp parallel corpus**

The article presents how the Treq corpus tool can be used in translation.

It also provides a description of Polish and Russian text resources found in the InterpCorp parallel corpus as well as Treq features.

Author uses samples of translation including single and multi-word units in order to show advantages and disadvantages of the Treq application.

Key words: Treq, InterCorp, translation equivalence, Czech, Polish, Russian