

Wyszukiwanie ekwiwalentów w Polsko-Słowackim Korpusie Równoległym

Searching for equivalents in the Polish-Slovak Parallel Corpus

Abstract

This paper deals with a comparison of two closely related languages, Polish and Slovak, and focuses on prefixed verbs, in the case of which this proximity is especially visible, and which may cause problems for translators and lexicographers. In the paper the prefixed verbs are treated as a lexicographic problem and a possible solution is presented that involves searching for equivalents in real translations.

A small parallel Polish-Slovak corpus was created for the purpose of this research. In the paper the process of compiling the corpus is described, starting with the acquisition of parallel texts, through texts processing and choosing corpus tools, and ending with the annotation and lemmatization of texts.

Next, an analysis of the equivalents of prefixed verbs found in the corpus was carried out to measure their lexicographic potential, i.e. their suitability to be included in the dictionary, based on their accuracy and frequency in different contexts and with different arguments. Four different levels of lexicographic potential are distinguished: high, average, low and zero lexicographic potential.

The paper presents the preliminary results of corpus analysis of the lexicographic potential of the Slovak equivalents of Polish prefixed verbs which focuses on suitable lexicographic material. Since this paper presents a small part of more extensive research on verbs with different prefixes, only an analysis of equivalents of verbs with prefix *u-* and *roz-* (*ubierać*, *ubrać*, *ukrywać*, *ubolewać*, *rozciągając się*, *rozleć się*) is presented.

The last part of the paper discusses problems found along the way, and it considers the adjustments to the evaluation process. The problems include, among others: the size of the corpus, the number of occurrences, the evaluation method as such (where the goal is to adjust it so as it is the most objective method possible but without compromising the intuitions of native or proficient speakers of the two languages), and, finally, the inconsistency of the information provided

by the Word Sketch tool in the Sketch Engine, which produces slightly different results for the Polish and Slovak reference corpora.

Overall, the presented analysis demonstrates how parallel corpora can be used to improve the quality of bilingual dictionaries by providing translation equivalents from the real translations.

Słowa kluczowe: korpus równoległy, język polski, język słowacki, ekwiwalent, czasowniki przedrostkowe, leksykografia dwujęzyczna, transfer językowy

Keywords: parallel corpus, Polish, Slovak, equivalent, prefixed verbs, bilingual lexicography, linguistic interference

1. Wstęp

Polsko-Słowacki Korpus Równoległy został zbudowany w celu stworzenia bazy materiałowej do badań nad słowackimi ekwiwalentami tłumaczeniowymi polskich czasowników przedrostkowych.

Między językiem polskim i słowackim, jako bliskimi językami, często zachodzi transfer językowy rozumiany jako naruszenie normy językowej pod wpływem elementów z drugiego języka, polegający najczęściej na błędnym przyjmowaniu struktur gramatycznych z jednego języka do drugiego (Marićová, 2008: 93).

Grupa słów, w której zarejestrowaliśmy wysoką częstotliwość występowania negatywnego transferu językowego, to czasowniki przedrostkowe. Ich słowotwórstwo w obu językach jest formalnie identyczne (w obu językach poprzez dodawanie przedrostka do czasownika), zasób przedrostków jest formalnie prawie identyczny i to samo dotyczy też podstawowych czasowników, z których większość istnieje w obu językach tylko z małymi różnicami formalnymi. Wspominany transfer powoduje zmiany znaczeniowe czasowników przedrostkowych w obu językach. Transfer językowy z języka polskiego do języka słowackiego często polega na użyciu słowackiego przedrostka identycznego z polskim przedrostkiem, z tym że czasownik podstawowy jest tłumaczeniem polskiego czasownika podstawowego lub jego słowackim homonimem. Występowanie negatywnego transferu językowego zauważyliśmy zarówno w tłumaczeniach, jak i w pracach leksykograficznych (*uciąć – uťať, unosić – unášať*).

Uważamy tę kwestię nie tylko za problem tłumaczeniowy, ale także leksykograficzny. Naszym celem jest wyszukanie w utworzonym korpusie równoległym wystarczającej liczby ekwiwalentów przekładowych w celu poddania ich dalszej analizie pod kątem przydatności jako ekwiwalentów słownikowych.

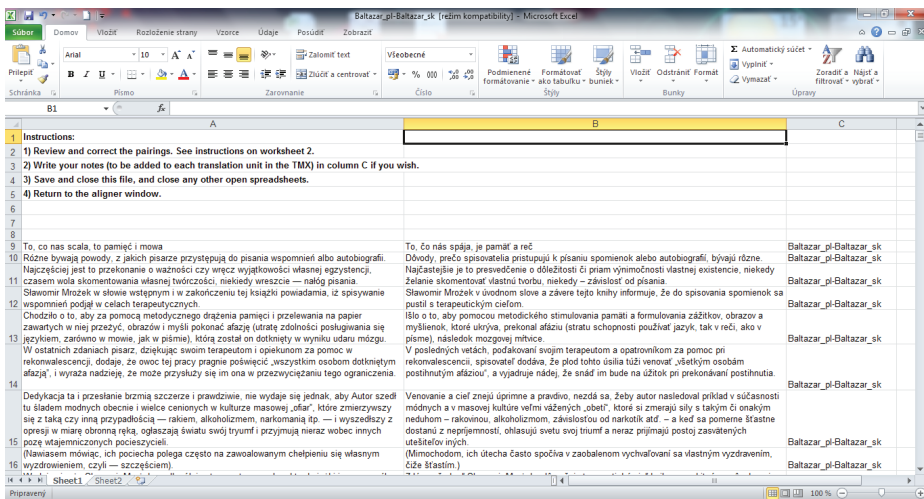
2. Utworzenie Polsko-Słowackiego Korpusu Równoległego

Korpus równoległy utworzony został z tekstów w języku polskim i słowackim. Korpus jest jednokierunkowy: językiem tekstów wyjściowych jest język polski, a językiem docelowym język słowacki. O jednokierunkowym charakterze korpusu zdecydował charakter i cel pracy, którym jest wyszukiwanie ekwiwalentów słowackich.

Wybór tekstów do korpusu został uwarunkowany możliwościami pozyskiwania tekstów równoległych w badanych językach, rodzajem tekstów, które są tłumaczone z polskiego na słowacki, ich dostępnością, a także pożądaną jakością tłumaczenia, która może stanowić problem ze względu na bliskość języków i spowodowane tym interferencje. Z wymienionych powodów do korpusu zostały wybrane teksty opublikowane w postaci książkowej, co znaczy, że poza tłumaczem w tłumaczeniu brała udział przynajmniej jedna dodatkowa osoba (redaktor, ewentualnie korektor lub weryfikator). Teksty do korpusu uzyskane zostały bezpośrednio od tłumaczy literatury polskiej (w ten sposób udało się pozyskać 10 książek z liczbą około 700 tysięcy słów – tokenów w części słowackiej). Następnie nawiązano kontakt z autorami tekstów polskich. Niestety, nie przyniosło to oczekiwanych rezultatów, gdyż udało się pozyskać tylko jedną książkę. Kilku autorów nie miało do dyspozycji ostatecznej wersji tekstu i poleciło kontakt z wydawnictwem. W niektórych przypadkach w ogóle nie udało się nawiązać kontaktu z autorem. W pozostałych wypadkach kontaktowaliśmy się z wydawnictwami (Znak, W.A.B, Noir sur Blanc i Czarne), w których, z wyjątkiem dwóch książek, udało się pozyskać wszystkie. Całkowita liczba słów w tekstach słowackich, dla których pozyskaliśmy odpowiedniki polskie, wynosi około 610 tysięcy. Wśród tekstów książek, które ostatecznie znalazły się w korpusie, znajdują się powieści (4 książki), eseje (2 książki), literatura faktu (1 książka) i literatura dla dzieci (1 książka). Nieco później natknęliśmy się na polskie czasopismo *Miłujcie się!* wydawane w różnych wersjach językowych, także w języku słowackim. Redakcja była chętna udostępnić nam teksty polskie i słowackie dwudziestu pięciu numerów czasopisma. Jednak ze względu na objętość już istniejącego korpusu oraz na to, że teksty ze wspomnianego czasopisma wymagałyby poszukiwań odpowiadających sobie tekstów w poszczególnych językach (tj. tekstów wyjściowych i docelowych) na razie zrezygnowano z ich włączenia do korpusu.

Oryginalne teksty otrzymaliśmy częściowo w formacie doc, a w większej części w formacie pdf. Te ostatnie trzeba było zapisać jako pliki doc i formalnie ujednolicić z tłumaczeniem. W plikach konwertowanych z formatu pdf potrzebna była kontrola znaków diakrytycznych, a także usunięcie znaków, które nie były częścią samego tekstu. Następnie teksty zostały zapisane w formacie

txt z kodowaniem UTF-8 i zrównoleglone na poziome zdań za pomocą oprogramowania LF_aligner opartego na oprogramowaniu hunalign, który służy do obsługi formatów takich jak txt, doc, docx, rtf, html, pdf z formatami wyjściowymi – tmx, txt i xls. W przypadku polsko-słowackiego korpusu plik wyjściowy był w formacie xls, ze zdaniami tekstu polskiego w pierwszej kolumnie i zdaniami słowackiego tłumaczenia w drugiej.



Rysunek1. Zrównoleglony plik xls

Corpora

Corpus name	Language	Tokens	Words
British National Corpus	English	112,181,015	96,048,950
ukWaC	English	1,565,274,190	1,318,612,719
BIBLE Polish_p1ah11-Polish	Polish	169,934	138,216
OPUS2 Polish	Polish	285,188,755	208,008,636
Polish Web Corpus v1.0 old	Polish	128,223,738	102,995,606
Polish Web Corpus	Polish	128,185,119	103,028,410
skTenTen	Slovak	876,003,720	715,707,053
esTenTen11 (Cu + Am, Freeing) sample	Spanish	56,913,874	48,721,039

[Show more corpora Parallel corpora](#)

My corpora

Corpus ID	Corpus name	Language	Configuration template	Tokens
english_skling	english_skling	English	Tree Tagger for English	257,979
babula_pl	babula_pl	Polish	Word, tag, lemma	444,129
babula_sk	babula_sk	Slovak	babula_sk_template	438,178
kivik	kivik	Slovak	General tokeniser	109,797

Rysunek 2. Sketch Engine z korpusami własnymi (powyżej) i korpusami użytkownika (poniżej)

Zrównoleglenie za pomocą oprogramowania LF_aligner jest bardzo skuteczne. Podczas ręcznego sprawdzania znaleziono tylko kilka błędów spowodowanych przede wszystkim niestandardowymi znakami na końcach i początkach

zdań (cudzysłowy, trzy kropki itp.), ewentualnie opuszczeniami całych zdań dokonanych przez tłumaczy. W tym przypadku polskie zdania, które znalazły się bez tłumaczenia słowackiego, zostały usunięte.

Jako oprogramowanie korpusowe wybraliśmy komercyjne oprogramowanie online Sketch Engine, w którym można tworzyć własne korpusy, w tym także korpusy równoległe.

Support

Help
Support

File preview (vertical)

|< << Showing lines 176..200 / 301892 >> >|

#	word	tag	lemma
176	stanowisko	stanowisko	subst:sg:acc:n
177	w	w	prep:loc:nwok
178	sprawach	sprawa	subst:pl:loc:f
179	dnia	dzień	subst:sg:gen:m3
180	dzisiejszego	dzisiejszy	adj:sg:gen:m3:pos
181	i	i	conj
182	w	w	prep:loc:nwok
183	sprawach	sprawa	subst:pl:loc:f
184	pamięci	pamięć	subst:sg:gen:f
185	<g/>		
186	.	.	interp
187	</align>		
188	<align>		
189	Ja	ja	ppron12:sg:nom:m1:pr1
190	jestem	być	fin:sg:pr1:imperf
191	tylko	tylko	conj
192	<g/>		
193	,	,	interp
194	by	by	comp
195	tak	tak	adv:pos
196	rzec	rzec	inf:perf
197	<g/>		
198	,	,	interp
199	logo	logo	subst:sg:nom:n
200	'	'	interp

Rysunek 3. Oznakowany i zlematyzowany plik wertykalny.

3. Wyszukiwanie w korpusie równoległym

Do wyszukiwania planowaliśmy wykorzystać wyrażenia regularne, dlatego należało tekst oznakować i lematyzować. Oznakowanie i lematyzowanie zostało wykonane w Centrum przetwarzania języka naturalnego na Wydziale Informatyki Uniwersytetu Masaryka w Brnie. Rezultatem są pliki wertykalne, które można zapisać do Sketch Engine i w ten sposób otrzymać oznakowany i lematyzowany korpus.

Do wyszukiwania zastosowano wyrażenia regularne w formie: [lemma="u.*" & tag="inf.*|fin.*|praet.*|imps.*|pact.*|ppas.*|pcon.*|impt.*"], co znaczy, że wyszukiwane zostały słowa z przedrostkiem *u-*, morfologicznie oznakowane jako bezokolicznik, czasownik w czasie teraźniejszym, czasownik w czasie przeszłym, w formie bezosobowej czasu przeszłego, imiesłów przymiotnikowy czynny i bierny, imiesłów przysłówkowy i czasownik w trybie rozkazującym. Liczba znalezionych pojedynczych form jest bardzo wysoka (tysiące wystąpień dla większości przedrostków), należało więc zdecydować, które formy będą dalej analizowane. Na wstępie z wyrażenia regularnego usunięte zostały formy imiesłów i jego obecna forma, tj.: [lemma="u.*" & tag="inf.*|fin.*|praet.*|imps.*|impt.*"]. Część mowy ekwiwalentu, jak i jego częstotliwość w tekście, też prawdopodobnie będą czynnikami decydującymi, ponieważ określają jego zdolność do pełnienia roli ekwiwalentu słownikowego.

Do tej pory wykonano wyszukiwania i analizę czasowników z przedrostkami *u-*, *roz-* i *do-*. Rezultaty pojedynczych wyszukiwań były eksportowane w formie pliku tekstowego. Następnie wykonano selekcję par ekwiwalentowych do analizy. Ze znalezionych par usunięto te, w których ekwiwalenty są identyczne z ekwiwalentami znajdującymi się w słowniku polsko-słowackim, ponieważ celem analizy jest znalezienie nowych ekwiwalentów. Zostały również usunięte czasowniki polskie z kombinacją liter *roz-*, *do-* lub z literą *u-* na początku, w których nie pełnią one roli przedrostka (np. w czasownikach *rozumować*, *rozumieć* itp.).

4. Potencjał leksykograficzny

Ponieważ, jak wspomniano, celem analizy jest znalezienie słowackich ekwiwalentów leksykograficznych dla polskich czasowników przedrostkowych, wykonano ocenę zdolności ekwiwalentów pochodzących z korpusu równoległego do pełnienia roli ekwiwalentu słownikowego. Ta zdolność była mierzona jako potencjał leksykograficzny. Termin ten został zaczerpnięty od Magdaleny Perdek (Perdek, 2011: 262), która analizuje polskie ekwiwalenty angielskich czasowników frazowych. Definiuje ona potencjał leksykograficzny jako zdolność ekwiwalentu do włączenia go do słownika. Zdolność ta opiera się na dokładności i zdolności do zastosowania w różnych kontekstach i z różnymi argumentami.

Perdek rozróżnia cztery stopnie potencjału leksykograficznego (Perdek, 2011: 262-263):

1. Wysoki PL – synonimiczny z ekwiwalentem leksykograficznym i z bardzo wysokim zastosowaniem w różnych kontekstach i z najczęstszej występującymi argumentami.

2. Średni PL – semantycznie podobny do ekwiwalentu leksykograficznego, ale ze względu na różnice w strukturze i odmienne kolokacje, ma ograniczony zasięg.
3. Niski PL – ekwiwalent korpusowy jest tłumaczeniem wykładu, co ogranicza konteksty i argumenty, w których można go zastosować i czasami rozszerza lub zawęża jego oryginalne znaczenie.
4. Zerowy PL – ekwiwalent korpusowy jest ograniczony tylko do jednego kontekstu i nie jest możliwe jego rozszerzenie na więcej kontekstów lub argumentów. Do tej grupy należą też ekwiwalenty, które są rezultatem błędnego tłumaczenia.

5. Metoda oceny

Ocenę wykonywano w postaci identyfikacji znaczenia, w którym zastosowano polski czasownik i jego tłumaczenie. Jako materiał referencyjny zastosowano *Słownik języka polskiego* (<http://sjp.pwn.pl>) i słowackie słowniki dostępne online pod adresem (slovník.juls.savba.sk). Następnie zostało ocenione i porównane „zachowanie” kolokacyjne polskiego i słowackiego czasownika w korpusach referencyjnych (plTenTen12 i skTenTen11 – <https://the.sketchengine.co.uk>) za pomocą narzędzia Word Sketch, które generuje automatyczny przegląd gramatycznego i kolokacyjnego „zachowywania” się słowa w tekstach. Jako leksykograficzny materiał referencyjny zastosowano *Polsko-slovenský a slovensko-poľský slovník* (Stano, Buffa, 1975).

6. Analizy ekwiwalentów tłumaczeniowych

W następnej części przedstawimy kilka przykładów analizowanych par czasownikowych.

ubierać	zdobiť	Wysoki
I kiedy w Boże Narodzenie ubieraliśmy drugą choinkę	A keď sme na Božie narodenie zdobili ďalšiu jedličku – prvú	Obliekať
– pierwszą u siebie w domu, drugą w Oliwie...	u nás doma a druhú v Oliwie...	

Tabela 1. Ekwiwalent tłumaczeniowy czasownika *ubierać*

Czasownik *ubierać* użyty jest w znaczeniu: ‘ozdobić, przystroić’ (*Słownik języka polskiego* [s.a.]). Hasło w polsko-słowackim słowniku ma tylko jedno znaczenie i tylko jeden ekwiwalent: *obliekať* (Stano, Buffa, 1975: 354). Według narzędzia Word Sketch jedynym przedmiotem dla tego znaczenia jest właśnie rzeczownik

choinka. Słowackiego czasownika *zrobiť* używa się z rzeczownikami, które są nazwami przedmiotów, takimi jak *stena*, *fasáda*, *priečelie*, *interiér*, ale też *stromček*, *kraslica*. Ekwiwalent w tym znaczeniu ma szersze zastosowanie, ale możemy mu przypisać wysoki potencjał leksykograficzny.

ubrać	zahaliť	zerowy
Wrócił fagas, a za nim weszły dwie kobiety ubrane do kostek i w chustkach.	Vrátil sa lokaj a za ním vošli dve ženy zahalené v šatách po členky a v šatkách.	obliecť

Tabela 2. Ekwiwalent tłumaczeniowy czasownika *ubrać*

Polski czasownik *ubrać* użyty jest w znaczeniu: ‘włożyć na kogoś ubranie, obuwie, nakrycie głowy i p.’ (*Słownik języka polskiego* [s.a.]). Jak wspomnieliśmy wcześniej, hasło w słowniku polsko-słowackim ma tylko jeden ekwiwalent *obliecť* (Stano, Buffa, 1975: 354). Ekwiwalent *zahaliť* (okryć) jest motywowany kontekstem *do kostek i w chustkach* i dobrze określa wizualny obraz, który się nasuwa po przeczytaniu zdania. W innych zwykłych użyciach czasownika *ubrać* – nie użylibyśmy tego ekwiwalentu, jeżeli nie byłby spełniony warunek odpowiedniego kontekstu.

ukrywać	tajiť	średni
Partia, bolszewicki wynalazek wcale przez wynalazców nieukrywany , nie była w moim guście.	Strana, bolševický vynález, ktorým sa jej vynálezcovia vôbec netajili , absolútne nebola v mojom štýle.	ukryvať, skrývať

Tabela 3. Ekwiwalent tłumaczeniowy czasownika *ukrywać*

Polski czasownik *ukrywać* użyty jest w znaczeniu: ‘nie ujawnić swoich myśli, uczuć, zamiarów’ (*Słownik języka polskiego* [s.a.]). Hasło w polsko-słowackim słowniku ma tylko jedno znaczenie z ekwiwalentami *ukryvať*, *skrývať* (Stano, Buffa, 1975: 357). Polskiego czasownika używa się często z rzeczownikami będącymi nazwami emocji np: *radość*, *satysfakcja*, *duma*, *wzruszenie*, *pogarda*, *zazdrość*, *zadowolenie*, *zachwył*, ale też z innymi, takimi jak *tożsamość* i *orientacja*. Słowackiego ekwiwalentu, według narzędzia Word Sketch, używa się częściej z rzeczownikami takimi jak *dych*, *nevera*, *tehotenstvo*, *orientácia*, *identita*, a rzadziej z przedmiotami oznaczającymi uczucia. Tych ostatnich używa się raczej z czasownikiem *skrývať*.

ubolewać	žialiť	wysoki
Walentin Rasputin, prawdziwy wielkoruski patriota, skądinąd autor znakomitej prozy, ubolewa nad wrogością świata wobec patriotyzmu rosyjskiego.	Valentin Rasputin, skutočný veľkoruský vlastenec, inak autor vynikajúcej prózy, žiali nad nepriateľstvom sveta voči ruskému vlastenectvu.	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)

Tabela 4. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

Czasownik *ubolewać* użyty jest w znaczeniu: ‘odczuwać i wyrażać żal, współczucie lub niezadowolenie’ (*Słownik języka polskiego* [s.a.]). W haśle w polsko-słowackim słowniku przytoczone są ekwiwalenty *ľutovať* (koho, čo), *nariekať* (nad kým, čím) (Stano, Buffa, 1975: 354). Według narzędzia Word Sketch czasownika *ubolewać* najczęściej używa się z przyimkiem *nad* i łączy się z rzeczownikami: *fakt, brak, los, strata, upadek*, czyli ze słowami, które zawierają semantyczną cechę wartości negatywnej. Słowacki czasownik *žialiť* łączy się z kilkoma przyimkami, z których najczęściej występuje przyimek *nad*, łączący się z słowami takimi jak *strata, smrť, odchod, osud, neúspech*. Znów są to więc słowa z semantyczną cechą wartości negatywnej lub takie, które mogą taką cechę nabyć w kontekście.

ubolewać	vyslovovať ľútosť	niski
Czytam emfaticzny wstęp Károla Kincsesa, który	Čítam emfatický úvod Károla Kincsesa, ktorý	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)
ubolewa nad tym, że żaden z fotografów, którzy zostali w kraju, nie zrobił światowej kariery.	vyslovuje ľútosť nad tým, že nikto spomedzi fotografov, čo zostali doma, neurobil svetovú kariéru.	

Tabela 5. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

W tym przypadku słowackiego ekwiwalentu używa się podobnie jak polskiego czasownika, ale zgoda znaczeniowa jest tylko częściowa, ponieważ czasownik słowacki zawiera jedynie komponent ‘wyrażać’ natomiast nie zawiera komponentu ‘odczuwać’ (*Słownik języka polskiego* [s.a.]). Ze względu na tę różnicę oraz dlatego, że ekwiwalentem jest wyrażenie składające się z kilka słów, przyznajemy mu niski potencjał leksykograficzny.

ubolewać	smútiť	wysoki
Odpowiadam: Myślę, że sądy są wciąż niezawisłe, nad czym najwyraźniej ubolewają premier Jarosław Kaczyński i minister sprawiedliwości Zbigniew Ziobro.	Odpovedám: Myslím si, že súdy sú stále nezávislé, nad čím evidentne smútia premiér Jarosław Kaczyński a minister spravodlivosti Zbigniew Ziobro.	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)

Tabela 6. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

Ekwiwalentu *smútiť*, podobnie jak *žialiť*, używa się z formami po przyimkach. Według narzędzia Word Sketch najczęstszym przyimkiem jest *za*, a drugim *nad*. Rzeczowniki łączące się z tymi przyimkami to przede wszystkim *strata, smrť, minulosť, osud, láska, tragédia, život*. Sytuacja jest podobna jak przy ekwiwalencie *žialiť*, natomiast na korzyść ekwiwalentu *smútiť* przemawia o wiele wyższa częstotliwość. W korpusie skTenTen11 jest ona prawie sześć razy wyższa.

ubolewać	lamentovať	średni
Oto przed kilku laty redaktor jednego z pism katolickich o kierunku konserwatywnym, ubolewając , że nie przeprowadzono szczegółowych badań dotyczących udziału Żydów w aparacie władzy komunistycznej, wypowiedział pogląd wart zacytowania:	Například pred niekoľkými rokmi redaktor jedného z katolíckych konzervatívne zameraných časopisov, lamentujúc nad tým, že sa neuskutočnili detailné vyšetrovania týkajúce sa podielu Židov v aparáte komunistickej moci, vyslovil názor hodný citovania:	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)

Tabela 7. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

Ekwiwalent *lamentovať* najczęściej występuje razem z formami po przyimkach z wyraźną przewagą przyimka *nad*, który się łączy z rzeczownikami o negatywnym nacechowaniu lub takimi, które je mogą nabyć pod wpływem kontekstu, jak na przykład *zdravotný stav, následky, situácia, nešťastie, život, osud, minulosť*. Czasownik *lamentovať* w języku słowackim jest rzadszy i również mocniej nacechowany niż polskie *ubolewać*. To potwierdza też *Krátky slovník slovenského jazyka*: „expr. nahlas prejavovať žiaľ, nespokojnosť, bolesť ap., horekovať, bedákať, nariekať” (Kačala, Pisárčiková, Považaj, 2003). Ze względu na różnicę w nacechowaniu jego potencjał leksykograficzny oceniono jako średni.

ubolewać	žalostiť	średni
Dowiaduję się o artykułach w węgierskiej skrajnie prawicowej prasie, których autorzy ubolewali nad tym, że Węgry nie wykorzystały wojen bałkańskich do odzyskania przynajmniej części dawnych ziem.	Dozvedám sa o článkoch v maďarskej krajne pravicovej tlači, ktorých autori žalostia nad tým, že Maďari nevyužili balkánske vojny na znovuzískanie aspoň časti stratených území.	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)

Tabela 8. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

Ekwiwalent *žalostiť* ma ze wszystkich przytoczonych ekwiwalentów czasownika *ubolewać* najmniej wystąpień. Najczęściej występował w konstrukcjach z przyimkami, przede wszystkim z przyimkiem *nad* z wyrazami takimi jak: *osud, sa (sebou), ty (tebou), nespravodlivosť*. Czasownik *žalostiť* ma w słowniku *Krátky slovník slovenského jazyka* kwalifikator stylistyczny „niż.” (książk.) i jego definicję słownikową wyrażają synonimy: „žiaľiť, smútiť, nariekať, bedákať” (Kačala, Pisárčiková, Považaj, 2003). Ze względu na wyższe nacechowanie stylistyczne jego potencjał leksykograficzny oceniono jako średni.

ubolewać	bedákať	średni
Ale jeżeli należy ubolewać nad tym, że na froncie politycznym w Polsce giną Żydzi, to trzeba ubolewać również nad tym, że giną, i to w znacznie większej ilości, Polacy.	Ale ak treba bedákať nad tým, že na politickom fronte v Polsku sa strácajú Židia, takisto treba bedákať aj nad tým, že miznú, a to v oveľa väčšom množstve, Poliaci.	ľutovať (niekoho, niečo), nariekať (nad niekým, niečím)

Tabela 9. Ekwiwalent tłumaczeniowy czasownika *ubolewać*

Ekwiwalent *bedákať*, podobnie jak inne tutaj przytoczone ekwiwalenty, najczęściej występuje w wyrażeniach przyimkowych z wyraźną przewagą przyimka *nad* z rzeczownikami: úbohosť, nespravodlivosť, zloba, osud, bieda, nešťastie. W porównaniu z polskim czasownikiem *ubolewać* czasownik słowacki *bedákať* zawiera komponent semantyczny nawiązujący do czynności płakania, co wpływa na różnicę między nimi.

rozciągać się	tiahnuť sa	średni
Widok z okna na zachód rozciągał się aż do ulicy Bandurskiego.	Výhľad z okna na západ sa tiahol až po Bandurského ulicu.	-

Tabela 10. Ekwiwalent tłumaczeniowy czasownika *rozciągać się*

Czasownik *rozciągać się* jest użyty w znaczeniu: 'zajmować pewną przestrzeń' (*Słownik języka polskiego* [s.a.]). Prototypowym ekwiwalentem dla tego znaczenia jest czasownik *rozkladać sa*. Najczęstszymi połączeniami polskiego czasownika *rozciągać się* są rzeczowniki: *widok*, *panorama*, *park*, *krajobraz*. Jeśli spojrzymy na słowacki czasownik *tiahnuť sa*, wśród rzeczowników, z którymi się łączy, nie znajdziemy rzeczownika *výhľad*, i tak samo odwrotnie (wśród czasowników używanych z rzeczownikiem *výhľad* nie znajdziemy czasownika *tiahnuť sa* (to samo potwierdzają wyszukiwania na www.google.sk). Natomiast z innymi wspomnianymi rzeczownikami czasownik *tiahnuť sa* może być używany. Są to zazwyczaj rzeczowniki, które oznaczają rodzaj krajobrazu naturalnego lub stworzonego przez człowieka – *pobrežie*, *pláž*, *pohorie*, *hrebeň*, *polostrov*, *múr*.

rozlec się	zazvoníť	niski
Nawiązując do stwierdzenia Winstona Churchilla, napisałeś, że jeśli w obecnej Polsce gdzieś o szóstej rano rozlegnie się dzwonek u drzwi, to nikt już nie może być pewny, że to tylko mleczarz.	V súvislosti s konštatovaním Winstona Churchilla si napísal, že ak v súčasnom Poľsku niekde o šiestej ráno zazvoní zvonček pri dverách, už si nikto nemôže byť istý, že je to len mliekar.	-

Tabela 11. Ekwiwalent tłumaczeniowy czasownika *rozlec się*

Czasownik *rozlec się* użyty jest w znaczeniu ‘dać się słyszeć’ (*Słownik języka polskiego* [s.a.]). Ekwiwalent *zazvonit’* jest tłumaczeniem czasownika razem ze znaczeniem rzeczownika, z którym został użyty, jest więc kontekstowo ograniczony do tych argumentów, które wyrażają dany dźwięk, przede wszystkim *dzwon* lub *dzwonek*.

rozlec się	rozlahnúť sa	średni
I tak trwaliśmy do północy w ogólnej niepewności, aż rozległ się ogłuszający huk, słyszalny nawet w piwnicy.	A tak sme čakali do polnoci v úplnej neistote, až sa rozlahol ohlušujúci huk, ktorý bolo počuť aj v pivnici.	-

Tabela 12. Ekwiwalent tłumaczeniowy czasownika *rozlec się*

Polski czasownik *rozlec się* jest tu użyty w tym samym znaczeniu jak w poprzednim przykładzie. Ekwiwalent *rozlahnúť sa* jest ograniczony do użycia z rzeczownikami, które oznaczają bardzo głośne dźwięki lub *ticho*.

7. Ocena fazy pilotażowej

W analizie przedstawionej w tej prezentacji pokazano, jak równoległe korpusy mogą być zastosowane do podniesienia jakości słowników dwujęzycznych dzięki dostarczeniu ekwiwalentów tłumaczeniowych wyekscerpowanych z przekładów. Ze względu na to, że przedstawione tu badania wciąż znajdują się na etapie początkowym, staramy się dostosowywać procedury oceny ekwiwalentów do pojawiających się problemów. Przede wszystkim należy wziąć pod uwagę wysoką liczbę znalezionych czasowników; zmieniono także formułę wyrażenia regularnego. Kolejnym krokiem będzie prawdopodobnie zastosowanie metody próbkowania. Istotnym problemem jest ciągle potrzeba dostosowywania metody oceny tak, żeby była ona jak najbardziej obiektywna, ale bez rezygnacji z wiedzy i świadomości rodzimego użytkownika badanych języków lub użytkownika na poziomie zaawansowanym. Wreszcie ostatnim z zauważonych problemów jest jakość i ilość informacji dostarczanych przez narzędzie Word Sketch, ponieważ wyniki, które są otrzymywane, mają umiarkowanie odrębny charakter dla polskiego i słowackiego korpusu referencyjnego.

Mimo problemów napotykanych w trakcie prowadzenia badań, mamy nadzieję, że otrzymany materiał będzie można wykorzystać do pracy nad przyszłymi słownikami polsko-słowackimi.

Literatura

- KAČALA, Ján, PISÁRČIKOVÁ, Mária, POVAŽAJ, Matej (2003): *Krátky slovník slovenského jazyka*. 4. vydanie. slovník.juls.savba.sk, (25-03-2015).
- MARIČOVÁ, Anna (2008): *Slovesné predpony v slovenčine a srbčine*. Báčsky Petrovec: Slovenské vydavateľstvo centrum.
- PERDEK, Magdalena (2011): *English phrasal verbs in translation: A lexicographic and corpus study of equivalence*. (rozprawa doktorska). Poznań: Uniwersytet im. Adama Mickiewicza.
- plTenTen12. Sketch Engine [s.a.]: <http://www.sketchengine.co.uk>, (25-03-2015).
- skTenTen11. Sketch Engine [s.a.]: <http://www.sketchengine.co.uk>, (25-03-2015).
- Słownik języka polskiego [s.a.]: sjp.pwn.pl, (25-03-2015).
- STANO, Mikuláš, BUFFA, Ferdinand (1975): *Polsko-slovenský a slovensko-poľský slovník*. –Bratislava: Slovenské pedagogické nakladateľstvo.

