PF 2015 (LXVII): 131–150

ELŻBIETA KACZMARSKA
Instytut Slawistyki Zachodniej i Południowej
Uniwersytet Warszawski
Krakowskie Przedmieście 26/28
02-678 Warszawa
tel.: +48 22 55 20 311
e-mail: e.h.kaczmarska@uw.edu.pl

W POSZUKIWANIU ZNACZENIA CZASOWNIKÓW WYRAŻAJĄCYCH STANY PSYCHICZNE. ANALIZA CZESKICH CZASOWNIKÓW I ICH POLSKICH EKWIWALENTÓW – PRÓBA IMPLEMENTACJI WYBRANYCH TEORII LINGWISTYCZNYCH

(WALENCJA, GRAMATYKA PRZYPADKÓW GŁĘBOKICH, PATTERN GRAMMAR, LINGWISTYKA KOGNITYWNA)

SŁOWA KLUCZOWE: ekwiwalent, czasowniki wyrażające stany psychiczne, walencja, korpus równoległy

KEY WORDS: equivalent, psych verbs, valency, parallel corpora

Celem artykułu jest prezentacja wybranych metod ustalania polskich ekwiwalentów czeskich czasowników wyrażających stany psychiczne. Czasowniki te, w odczuciu rodzimego użytkownika języka polskiego, są szczególnie problematyczne. Ich wieloznaczność sprawia duże trudności w procesie przekładu, zdarza się bowiem, iż osoba posługująca się językiem polskim jako ojczystym nie może odkodować znaczenia czeskiego wyrazu i w rezultacie staje w obliczu

niemożności wyrażenia tej samej treści we własnym języku[1]. Często w takich sytuacjach obserwujemy, iż język docelowy nie dysponuje identycznym pojęciem (Kaczmarska i Rosen 2013c). Z tego powodu podczas tłumaczenia tracimy część znaczenia lub nawet znaczeń. Możemy wprawdzie wymienić kilka ekwiwalentów dla każdej jednostki, ale są to ekwiwalenty defektywne. Lewandowska-Tomaszczyk stwierdza, że niemożliwym jest znalezienie w stu procentach trafnego ekwiwalentu dla jakiegoś słowa w języku obcym, ale możemy znaleźć cały klaster ekwiwalentów z odrobinę zmienionym znaczeniem (Lewandowska-Tomaszczyk 1984, 2013). Kwestię tłumaczenia może dodatkowo komplikować fakt, iż uczucia – jako pozbawione struktury – uważane są za niewyrażalne[2].

> „Uczucie to jest coś, co się czuje – a nie coś, co się przeżywa w słowach. W słowach można zapisać myśli – nie można zapisać w słowach uczuć. Myśl jest czymś, co ma strukturę dającą się odtworzyć słowami. Uczucie z natury rzeczy jest pozbawione struktury, a więc niewyrażalne" (Wierzbicka 1971, s. 30).

Proponowana analiza koncentruje się więc nie na możliwościach wyrażaniu uczuć w języku docelowym (tu – w polskim), a na poszukiwaniu polskich ekwiwalentów badanych czasowników użytych w konkretnym znaczeniu. Wspomnianych problemów nie rozwiązują tradycyjne słowniki, które z racji naturalnych ograniczeń oferują ograniczoną liczbę ekwiwalentów (w większości przypadków bez przykładów ilustrujących użycie czy kontekst). Najbardziej znany słownik czesko-polski (Siatkowski i Basaj 2002) prezentuje szereg ekwiwalentów dla danych czasowników, np.: *mrzet – gniewać, złościć, mierzić, martwić, żałować, być przykro, nie mieć ochoty*. Bez dodatkowych informacji nie jesteśmy w stanie przetłumaczyć ich poprawnie na język polski[3].

1. Analiza

Celem podjętego badania jest znalezienie odpowiedniego ekwiwalentu dla danej jednostki oraz – w rezultacie – próba zbudowania algorytmu wyszukującego

[1] Analiza nie obejmuje zjawiska „fałszywych przyjaciół", obecnego na styku polszczyzny i czeszczyzny, np.: (pl) *poprawić* – (cs) *popravit* (pl. *zgładzić, zamordować, uśmiercić*).

[2] Szerzej na temat sposobu wyrażania uczuć Pajdzińska 1990.

[3] W niektórych przypadkach rozstrzygający nie jest też kontekst – wyznanie *mám Tě rád* może być przetłumaczone jako *lubię cię* albo *kocham cię*. Dla osoby polskojęzycznej jednostka *mít rád* ma co najmniej dwa zupełnie różne znaczenia: *kochać* i *lubić*, zob. też: Kaczmarska i Rosen 2014a.

optymalny ekwiwalent dla tego typu czasowników[4]. Badanie dotyczy implementacji różnych metod językoznawczych[5]:

- automatyczna ekstrakcja par ekwiwalentów (Och i Ney 2003; Skoumalová 2008; Jirásek 2011) z korpusu równoległego InterCorp (Čermák i Rosen 2012),
- analiza pod względem wymogów walencyjnych (Daneš, Hlavsa 1987; Rytel 1989; Čermáková 2009; Urbańczyk-Adach 2011)[6],
- sprawdzanie zgodności przypadków semantycznych – Case Grammar (Fillmore 1968; Halliday 1985; Korytkowska 1984, 1992, 1993; Kaczmarska 2001, 2010),
- analiza wzorców, w jakich pojawiają się badane jednostki – Pattern Grammar (Francis i in. 1996; Hunston i Francis 2000; Ebeling i Ebeling 2013),
- wykorzystanie elementów gramatyki kognitywnej (Langacker 1987, 1991, 2008; Taylor 2002; Mikołajczuk 1997, 1999).

Badanie oparte jest na materiale z korpusu równoległego InterCorp (Čermák i Rosen 2012; Kaczmarska i Rosen 2014b). Tworzenie eksperymentalnego algorytmu będzie wymagało kilku kroków, odnoszących się do wspomnianych teorii gramatycznych. W zależności od badanej jednostki optymalny ekwiwalent można wskazać na każdym etapie analizy.

1.1. Krok pierwszy – automatyczna ekscerpcja par ekwiwalentów

Z korpusu równoległego InterCorp wyekstraktowane zostały pary ekwiwalentów, powstał dzięki temu swoisty dwujęzyczny słownik (Och i Ney 2003; Skoumalová 2008; Jirásek 2011; Kaczmarska i Rosen 2013b), zawierający oczywiście również błędnie dopasowane odpowiedniki (wynika to z niedoskonałego wyrównania segmentów). W poniższej tabeli (1) ukazano dziesięć najczęstszych ekwiwalentów dla czeskiego czasownika *mrzet*:

---

[4] Prace nad algorytmami ulepszającymi tłumaczenia maszynowe czy różnicującymi znaczenia jednostek wieloznacznych (np. WSD – *Word Sense Disambiguation*) są już na świecie prowadzone od dawna i znane także w odniesieniu do korpusów równoległych, w większości opierają się na olbrzymich danych uzyskanych z korpusów i bazują na różnych metodach matematycznych (w tym – statystycznych), np. Liang Tian i in. 2014; Młodzki i in. 2012; Liang Tian i in. 2010; Han i in. 2013; Kędzia i in. 2014. Opracowywane algorytmy wykorzystują też różne podejścia lingwistyczne. Szerzej na ten temat – Han i in. 2013.

[5] W artykule tym zostały opisane jedynie wybrane elementy kilku teorii lingwistycznych w odniesieniu do konkretnej grupy czasowników.

[6] W przyszłości ten etap analizy będzie uwzględniał również opis czasowników w programie RAMKI – konotacja, walencja, własności głębinowo-składnikowe (Linde-Usiekniewicz i in. 2008; Zawisławska 1998).

Tab. 1. Wyekstraktowane automatycznie odpowiedniki czeskiego czasownika *mrzet*. W lewej kolumnie zostały podane liczby bezwzględne poświadczeń w InterCorpie

Ekwiwalenty czasownika *mrzet*	
124	*przykro*
64	*żałować*
34	*przepraszać*
19	*żal*
11	*szkoda*
9	*martwić*
8	*pożałować*
8	*zmartwić*
6	*przykrość*
5	*gniewać*

Na tym etapie utworzono klastery polskich ekwiwalentów dla każdego z poszczególnych czeskich czasowników[7]. W kolejnej tabeli (2) zaprezentowano klaster najczęstszych ekwiwalentów czasownika *toužit*:

Tab. 2. Polskie ekwiwalenty czasownika *toužit*. W prawej kolumnie zostały podane liczby bezwzględne poświadczeń w InterCorpie

Ekwiwalenty czasownika *toužit*	
pragnąć	304
chcieć	107
tęsknić	82
marzyć	70
pożądać	40
ochota	24
zapragnąć	9
pragnienie	8
tęsknota	8
zależeć	8
spragniony	7
życzyć	6
W sumie	673

[7] Bogactwo ewentualnych ekwiwalentów wiąże się z faktem, iż każdy przekład jest swoistą interpretacją tłumacza, a wybrany i wykorzystany przez niego ekwiwalent to jedna z możliwych wersji. Tłumacz wybiera jeden odpowiednik z gamy jednostek synonimicznych (Greń i Rytel-Kuc 1991, s. 69–78; Kaczmarska 2001, s. 183).

Warto zwrócić uwagę na fakt, iż zestaw ten zawiera wszystkie trzy ekwiwalenty wskazane przez słownik tradycyjny (Siatkowski i Basaj 2002) – *tęsknić, marzyć, pragnąć*.

1.2. Krok drugi – analiza walencyjna

Celem analizy na tym poziomie badania jest odpowiedź, czy wymogi walencyjne[8] mogą ułatwić identyfikację polskiego ekwiwalentu danego czeskiego czasownika. W tym celu przeprowadzono badanie pilotażowe czasownika *toužit* (Kaczmarska i Rosen 2013b). Założono, iż w niektórych przypadkach ekwiwalent może być ustalony na podstawie zbieżności wymagań walencyjnych (Levin 1993). Kolejno przeprowadzano analizy manualne wybranych czasowników (wraz z wiązanymi segmentami). W każdym segmencie określano, jak dany czeski czasownik został przetłumaczony na język polski, ile argumentów wiąże i jakiego są one rodzaju[9] oraz w jaki sposób łączą się z danym czasownikiem[10]. Skontrolowano także, jakie argumenty łączą się z polskimi ekwiwalentami oraz przeprowadzono syntaktyczną i semantyczną analizę argumentów wiążących się z czeskimi czasownikami i ich polskimi ekwiwalentami. Oczekiwano, iż w większości przypadków wyniki analizy syntaktyczno-semantycznej pozwolą ustalić najbliższy pod względem składni i znaczenia polski ekwiwalent dla czeskiej jednostki.

Dzięki analizie manualnej wiązanych segmentów można było sprawdzić, jak badany czasownik został przetłumaczony w poszczególnych grupach. Okazało się, iż walencja może wpłynąć na wybór polskiego ekwiwalentu. W przypadku

---

[8] W artykule tym walencję rozumiemy jako zjawisko językowe odnoszące się do liczby argumentów łączących się z predykatem werbalnym (por. Daneš i Hlavsa 1987; Rytel 1989; Čermáková 2009; Urbańczyk-Adach 2011).

[9] Np. fraza nominalna – jeśli tak, to jakiego typu: obiekt realny, abstrakcyjny, nazywający istotę ludzką.

[10] Czasownik *toužit* wiąże argumenty, wykorzystując następujące wzorce:

– *toužit po* Oabstr (obiekt abstrakcyjny),

– *toužit po* Ohum (istota ludzka),

– *toužit po* / *do* OR (obiekt rzeczywisty),

– *toužit* + *inf* (konstrukcja z bezokolicznikiem).

– *toužit* + S (aby… / po tom, aby… – konstrukcje ze zdaniami podrzędnymi). Trzy pierwsze wzorce odnoszą się do syntaktycznie tej samej konstrukcji *toužit* + *po* + *O* (obiekt w miejscowniku). Rozróżnienie na trzy grupy dotyczy analizy semantycznej. Konieczność wyodrębnienia tych trzech grup zaobserwowano podczas syntaktycznej analizy manualnej. Tylko pozornie jest to pomieszanie dwóch płaszczyzn analizy. Praktycznie polegało to na dokładniejszym „tagowaniu" obiektu łączącego się z przyimkiem „po".

analizowanego czasownika optymalny ekwiwalent identyfikowany jest jednak tylko dla struktury *toužit* + bezokolicznik:

toužit + inf → pragnąć[11] *+ inf*.

Tab. 3. Polskie ekwiwalenty *toužit* + bezokolicznik

Ekwiwalent *toužit* + bezokolicznik	
pragnąć inf	44
chcieć inf	20
marzyć o Oabstr	4
pragnąć Oabstr	3
być pragnieniem inf	1
chętnie + S	1
mieć marzenie inf	1
mieć ochotę inf	1
pragnąć + S	1
tęsknić za (+S)	1
zachciewać się Oabstr	1
inne	2
W sumie	80

W pozostałych grupach uzyskane rezultaty nie są rozstrzygające[12]. Szczególnie struktury z obiektem abstrakcyjnym okazały się niezwykle problematyczne i wskazały na konieczność głębszej analizy semantycznej obiektów[13].

[11] Czasownik *chcieć* jest klasyfikowany jako synonim *pragnąć*. Uznajemy, iż różnica pomiędzy nimi polega na intensywności uczucia. Podobnie pozostałe jednostki, podkreślone w tabeli 3.

[12] Szczegółowe wyniki opublikowane zostały w Kaczmarska i Rosen 2013b.

[13] Przeprowadzono test z dwoma obiektami abstrakcyjnymi – *skutečné přátelství/prawdziwa przyjaźń*) i *nezapomenutelné dobrodružství/niezapomniana przygoda*. Zaobserwowano, iż oba obiekty łączą się z analizowanym czeskim czasownikiem *toužit po skutečném přátelství/po nezapomenutelném dobrodružství*. Podjęto więc próbę połączenia obu z trzema najczęstszymi ekwiwalentami polskimi. Okazało się, iż poprawne jest połączenie tych obiektów z czasownikiem *marzyć*:

(1) *Marzę o prawdziwej przyjaźni/niezapomnianej przygodzie*.

Z czasownikiem *tęsknić* może się łączyć obiekt *prawdziwa przyjaźń*, ale niepoprawne jest połączenie tego czasownika z frazą *niezapomniana przygoda*:

(2) *Tęsknię za prawdziwą przyjaźnią / niezapomnianą przygodą (???)*.

(3) *Tęsknię do prawdziwej przyjaźni / niezapomnianej przygody (???)*.

Również czasownik *pragnąć* dopuszcza połączenie z obiektem *prawdziwa przyjaźń*; połączenie tego czasownika z frazą *niezapomniana przygoda* jest natomiast dość niezręczne:

(4) *Pragnę prawdziwej przyjaźni/niezapomnianej przygody (?)*.

Test ten wykorzystywany był również do badania innych obiektów (Kaczmarska 2014; Kaczmarska, Rosen, Hana i Hladká, B. 2015).

Niejednoznaczność wyników badania na poziomie kroku pierwszego zmusza do bardziej szczegółowej analizy w kolejnym kroku. Wszystkie jednostki, które nie znalazły odpowiedniego ekwiwalentu, automatycznie przenoszone by były do dalszej analizy – następnego kroku w algorytmie.

1.3. Krok drugi – identyfikacja przypadków głębokich (Case Grammar)

Na tym poziomie ustalano, czy jednostki kandydujące na ekwiwalenty łączą się z obiektami reprezentującymi te same kategorie ról semantycznych. Poszukiwania idealnego ekwiwalentu może ułatwić identyfikacja obiektu o tej samej wartości, łączącego się zarówno z badanym czasownikiem czeskim, jak i potencjalnym ekwiwalentem polskim. Na początku warto jednak zauważyć, iż ze względu na swój charakter wszystkie analizowane czasowniki wyrażające stany psychiczne wiążą argumenty o wartości Experiencer[14], np.:

(5a) *Lucie mne miluje a její zdraví závisí na mé lásce! (kundera-zert)*[15].
(5b) *Łucja mnie kocha i jej zdrowie zależy od mojej miłości! (kundera-zert).*

(6a) *…nelekala se jeho kletby a celé armádě země i pekel se Satanem v čele by uměla odporovat svým sladce opovržlivým úsměvem?* (Durych-Bloudeni).
(6b) *…nie lękała się jego przekleństw i umiała całej armii ziemi i piekieł, z samym szatanem na czele, przeciwstawić swój słodko pogardliwy uśmiech.* (Durych-Bloudeni).

(7a) *Ale snad o to víc ho mrzelo, že petici nepodepsal* (Kundera-Nesnesit_lehko).
(7b) *Ale tym bardziej chyba żałował, że nie podpisał petycji* (Kundera-Nesnesit_lehko).

Występowanie obiektów o tej samej wartości jest tylko pozornym ułatwieniem, nie rozstrzyga bowiem kwestii wyboru ekwiwalentów. Warto się jednak przyjrzeć pozostałym obiektom obecnym w strukturze danych czasowników.

[14] Wyodrębnienie argumentu o wartości Experiencer wymagało ustalenia, jakie predykaty otwierają miejsce dla tego argumentu. Klasa tych predykatów została wyodrębniona na potrzeby tego opracowania na podstawie następujących kryteriów: 1) otwierają miejsce dla argumentu o cesze [+Hum] (wtórnie również dla argumentu o cesze [+Anim]); 2) odnoszą się do stanów psychosomatycznych i zdarzeń, które nie podlegają obiektywnej obserwacji, ponieważ dotyczą sfery wewnętrznej jednostki; 3) określają stany i zdarzenia zachodzące bez udziału woli jednostki, która jest nimi objęta; oznaczają stany i zdarzenia (nie czynności!), które jeśli wymagają pojawienia się „sprawcy zdarzenia", to jest to „sprawca" niedziałający świadomie i celowo. Por. Experiencer (Kaczmarska 2001, 2010; Korytkowska 1992, 1993), Proživatel (Karlík, Nekula, Rusínová 2002), Senser (Halliday 1985).

[15] Przykłady pochodzą z korpusu InterCorp; informacja bibliograficzna umieszczona jest na końcu każdej frazy w nawiasie.

Ponownie analizie poddano czeski czasownik *toužit*. W przykładach (8a)–(10b) podkreślone zostały pozostałe argumenty łączące się z badanym czasownikiem, ale niebędące obiektem o wartości Experiencer.

(8a) *Stál jsem i nyní stále kus od ní, kdežto ona naopak toužila po rychlém příchodu teplých doteků, které by přikryly tělo vystavené chladnosti pohledu* (kundera-zert).

(8b) *I teraz stałem nieco z dala od niej, podczas gdy ona, przeciwnie niż ja, tęskniła za szybkim dotknięciem ciepłych ramion, które osłoniłyby jej ciało wystawione na chłód spojrzeń* (kundera-zert).

(9a) *Toužil po polibku, závěrečném, posledním polibku, do kterého by zachytil jako do čeřenu její tvář, která brzy zmizí a z níž mu zůstane jen vzpomínka* (Kundera-Nesmrtelnost).

(9b) *Pragnął pocałunku, ostatniego pocałunku, kończącego pocałunku, który pozwolił by mu pochwycić niczym w sieć tę twarz, co wkrótce zniknie i pozostawi po sobie jedynie wspomnienie* (Kundera-Nesmrtelnost).

(10a) *Mladý muž touží po vlastním divadle* (Denemarkova-Mlady_muz).

(10b) *Młody mężczyzna marzył o własnym teatrze* (Denemarkova-Mlady_muz).

Jak już wcześniej wspomniano, partycypantem pojawiającym się w każdym z analizowanych przypadków jest argument o wartości Experiencera. Drugim argumentem występującym przy czasowniku *toužit* jest pewnego rodzaju źródło lub bodziec (impuls). W przypadku innych czasowników (lub również tego samego, ale w innej strukturze) zidentyfikować możemy także inne role, np. miejsce, czas, instrument, substancja, beneficjent (Location, Time, Instrument, Substance, Beneficiary). Nie zmieni to jednak charakteru jednostki wyrażającej stany psychiczne – w jego strukturze musi się znaleźć istota przeżywająca dany stan (tu – Experiencer) i bodziec ten stan wywołujący. Bodziec (źródło) nie musi się znajdować w tym samym zdaniu, może go ujawnić szerszy kontekst, ale dla zaistnienia stanu psychicznego jest konieczny:

(11a) *Podařilo se učinit infantu vzácnou; podmínky císaři byly kruté, ale čím byly krutější, tím více prý toužil mladý král* (Durych-Bloudeni).

(11b) *…Warunki postawione cesarzowi były okrutne, ale im okrutniejsze, tym mocniej podobno młody król pragnął infantki* (Durych-Bloudeni).

Ustalenie ról semantycznych argumentów łączących się z badanymi czasownikami i ich ekwiwalentami nie zbliża nas do wyłonienia adekwatnego ekwiwalentu. Wykorzystanie jako ogniwa algorytmu gramatyki przypadków głębokich może się okazać niewystarczające, ukazuje bowiem, iż powinna być przeprowadzona dokładna analiza powierzchniowej realizacji danych argumentów.

Ponownie pojawia się przypuszczenie, iż o wyborze ekwiwalentu może decydować leksykalno-semantyczna natura obiektów, czyli rodzaj otoczenia. Dotyczyć jej będzie następny poziom proponowanego algorytmu.

1.4. Krok trzeci – wzorce (Pattern Grammar)

Ten poziom analizy w założeniu ma pozwolić sprawdzić, czy istnieje zależność pomiędzy znaczeniem wyrazu a otoczeniem, w jakim występuje. Hunston i Francis twierdzą, że jeżeli jakieś słowo jest wieloznaczne, a jednocześnie pojawia się w kilku wzorcach, to każdy wzorzec pojawi się z jednym z jego znaczeń częściej niż z pozostałymi, czyli dany wzorzec wskaże najbardziej prawdopodobne znaczenie słowa.

> „If a word has several senses, and is used in several patterns, each pattern will occur more frequently with one of the senses than the others, such that the patterning of an individual example will indicate the most likely sense of the word in that example" (Hunston i Francis 2000, s. 20).

Czasowniki wyrażające stany psychiczne są w większości polisemiczne. Dzięki śledzeniu i identyfikacji ich wzorców może ułatwić połączenie konkretnych znaczeń z typem wzorca (rozumianym jako powtarzalna kombinacja słów)[16].

Podjęto próbę ustalenia, czy w rzeczywistości istnieje taka powtarzalność w poświadczeniach korpusowych (por. Ebeling i Ebeling 2013). Ręczna analiza oparta na materiale z korpusu InterCorp ukazała m.in. dwa wzorce czeskiej jednostki *být líto* (*być przykro, żałować*), związane z dwoma znaczeniami. Jeżeli jednostka *být líto* łączy się z dwoma frazami nominalnymi (w celowniku i dopełniaczu), wówczas taki jej wzorzec odpowiada polskiemu ekwiwalentowi *żal*. Jeżeli natomiast *být líto* łączy się tylko z celownikową frazą nominalną (oraz ewentualnie z elementem *to*), wzorzec ten odpowiada znaczeniu polskiej jednostki *(być) przykro*. Zostało to ukazane w poniższej tabeli (4).

Tab. 4. Wzorce czeskiej jednostki *být líto* (*żal, być przykro*)

żal	***(być) przykro***
(12a) *Jak mi ho bylo líto!*	(15a) *Pak mi je líto.*
(12b) *Jakże mi go było żal!*	(15b) *Wobec tego, przykro mi!*
(13a) *Je mi ho samozřejmě líto.*	(16a) *Potom nám to bylo oběma líto.*
(13b) *Jest mi go oczywiście żal…*	(16b) *Potem nam obu było przykro.*

[16] „A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it" (Hunston i Francis 2000, s. 37).

żal	*(być) przykro*
(14a) *Přišlo mi jí prostě líto.*	(17a) *…nabídne mi sisinku a já si vezmu, protože by mu bylo líto, kdybych si nevzala…*
(14b) *Po prostu zrobiło mi się jej żal.*	(17b) *…zaprasza mnie na cuksa i ja biorę, bo byłoby mu przykro, gdybym nie wzięła…*
být líto* + NP $_{DAT}$ + NP $_{GEN}$ = *żal	***být líto* + NP $_{DAT}$ + to / Ø = *(być) przykro***

Analiza manualna wzorców nie była niestety efektywna w przypadku czasownika mającego dużą liczbę poświadczeń (np. *toužit*). W wypadku takich jednostek warto użyć narzędzia Word Skeches.

1.5. Krok czwarty – Word Skeches

W przypadku Word Sketches nie chodzi o teorię językoznawczą. To program umożliwiający przedstawienie obrazu kolokacyjnego danego słowa również z odniesieniem do gramatyki[17]. Kolokaty analizowanej jednostki są pogrupowane według relacji gramatycznych, w których się pojawiają. Word Sketches wydają się więc uniwersalnym narzędziem do analizy kolokacji i połączeń wyrazowych rozumianych zarówno w kategoriach Pattern Grammar, jak również Case Grammar oraz walencji. Wykorzystując to narzędzie, dokonano próby ponownej analizy czeskiego czasownika *toužit*.

Badanie zostało przeprowadzone z wykorzystaniem czesko-polskiej części korpusu równoległego InterCorp. Na początku starano się potwierdzić tezę o ekwiwalencji *toužit* + bezokolicznik i *pragnąć* + bezokolicznik. Sprawdzono, który z ekwiwalentów łączy się z bezokolicznikiem. Materiał korpusowy ukazał, iż polskie czasowniki *marzyć* i *tęsknić* nie łączą się z bezokolicznikiem. Rezultaty zaprezentowano w tabeli 5., zawierającej jedynie dziesięć najczęstszych obiektów.

Tab. 5. Wyniki zastosowania narzędzia Word Sketches do analizy połączenia *toužit* + bezokolicznik i jego polskich ekwiwalentów

toužit inf		*pragnąć inf*		** marzyć inf*	** tęsknić inf*
post_inf	**17405**	**post_inf**	**6800**		
mít	*926*	*podziękować*	*805*		
stát	*864*	*podkreślić*	*598*		
poznat	*382*	*pogratulować*	*379*		
vidět	*346*	*wyrazić*	*391*		

[17] „A word sketch is a one-page, automatic, corpus-derived summary of a word's grammatical and collocational behavior" – http://www.sketchengine.co.uk/documentation/wiki/Website/Features#Wordsketches.

toužit inf		***pragnąć inf***		**** marzyć inf***	**** tęsknić inf***
vrátit	*333*	*zwrócić*	*319*		
hrát	*333*	*przypomnieć*	*165*		
získat	*332*	*powiedzieć*	*386*		
dostat	*311*	*zauważyć*	*97*		
vyhrát	*285*	*rozpocząć*	*73*		
žít	*177*	*poruszyć*	*58*		

W tym momencie pojawia się kolejny problem. Czeskie przykłady są ekscerpowane z olbrzymiego Czeskiego Korpusu Narodowego (dalej – CNK), a polskie przykłady – z korpusu równoległego InterCorp. Dane są nieporównywalne pod względem wielkości – por. poniżej wyniki ukazane w tabeli 6. Obecnie wykorzystanie narzędzi Word Sketches dla języka czeskiego jest niemożliwe z użyciem InterCorpu. Nie jest możliwe również wykorzystanie Narodowego Korpusu Języka Polskiego (dalej – NKJP). Wprawdzie jego wielkość jest porównywalna z wielkością czeskiego korpusu, ale NKJP nie oferuje narzędzia Word Sketches. Innym problemem są też różnice w rodzajach tekstów zamieszczonych w tych korpusach – Czeskim Korpusie Narodowym i czesko-polskiej części InterCorpu.

Tab. 6. Wyniki zastosowania narzędzia Word Sketches do analizy połączenia *toužit* + obiekt i jego polskich ekwiwalentów

toužit po		***pragnąć***		***marzyć o***		***tęsknić za***		***tęsknić do***	
post_po	**23752**	**has_gen_obj**	**809**	**verb_o_noun**	296	**verb_za_noun**	94	**verb_do_noun**	*59*
dítě	697	*co*	76	*to*	70	*dom*	6	*spokój*	3
láska	599	*to*	52	*Europa*	14	*to*	5	*dom*	3
návrat	555	*Europa*	34	*powrót*	10	*junior*	2	*świat*	3
úspěch	493	*zachęcić*	26	*demokracja*	5	*mąż*	2	*słońce*	2
vítězství	457	*strona*	16	*wolność*	4	*żona*	2	*ciało*	2
změna	455	*coś*	15	*utopia*	3	*powrót*	2	*rzecz*	2
život	361	*powód*	14	*zemsta*	3	*ojciec*	2		
medaile	316	*region*	14	*domek*	3	*coś*	2		
pomsta	287	*śmierć*	13	*kariera*	3	*człowiek*	2		
klid	282	*demokracja*	12	*miłość*	3	*czas*	2		

Wyniki w tabeli 6. są niejednoznaczne, postanowiono więc przeprowadzić eksperymentalne badanie z wykorzystaniem Word Sketches dla języka polskiego (InterCorp), kolokatora – również dla języka polskiego (NKJP)[18] oraz Word Sketches dla języka czeskiego (CNK). Wyniki tej analizy przedstawiono w tabeli 7.

[18] Kolokator (Pęzik 2012) dostępny na stronie: http://www.nkjp.uni.lodz.pl/collocations.jsp.

Tab. 7. Wyniki eksperymentalnego badania z użyciem kolokatora i Word Sketches

Word Sketches (InterCorp)		*kolokator*[19] *(NKJP)*		*Word Sketches (CNK)*	
pragnąć **has_gen_obj**	**809**	***pragnąć*** **+ Gen**		***toužit po*** **post_po**	**23752**
co	76	*on*	1460	*dítě*	697
to	52	*człowiek*	163	*láska*	599
Europa	34	*ty*	143	*návrat*	555
zachęcić	26	*życie*	110	*úspěch*	493
strona	16	*coś*	107	*vítězství*	457
coś	15	*bóg*	63	*změna*	455
powód	14	*kobieta*	60	*život*	361
region	14	*dziecko*	57	*medaile*	316
śmierć	13	*świat*	50	*pomsta*	287
demokracja	12	*nic*	47	*klid*	282

Rezultaty tego badania mają znaczenie wyłącznie poznawcze, a samego testu nie można implementować do algorytmu wyszukiwania ekwiwalentów. Program Word Sketches wydaje się obiecującym narzędziem dla analizy. Optymalną sytuacją byłoby, gdyby analiza została przeprowadzona na podstawie jednego korpusu i jednobrzmiących tekstów. Dlatego narzędzie, nad którym aktualnie trwają prace, współpracować będzie zarówno z czeską, jak i polską częścią InterCorpu. Powinno to umożliwić analizę nie tylko obiektów wyrażonych frazami nominalnymi, ale także przysłówków łączących się z danym czasownikiem. W poniższej tabeli 8. przedstawiono rozkład przysłówków w odniesieniu do interesujących dla badania jednostek[20].

Tab. 8. Łączliwość przysłówków – rezultat badania z wykorzystaniem Word Sketches

toužit		***pragnąć***		***tęsknić***		***marzyć***	
hodně	1099	*bardzo*	136	*bardzo*	34	*jedynie*	6
moc	1085	*gorąco*	40	*ogromnie*	3	*często*	5
tak	783	*jedynie*	19	*niesamowicie*	2	*bardzo*	5
už	778	*jednocześnie*	16	*okropnie*	2	*próżno*	4
tolik	773	*rozpaczliwie*	15	*straszliwie*	2	*długo*	4

[19] W przypadku kolokatora wyniki zostały zliczone ręcznie, a takiej analizie został poddany tylko jeden polski ekwiwalent – pragnąć.

[20] Ponownie jest to jednak badanie testowe, a więc wyników nie można brać pod uwagę we wnioskach czy podsumowaniu, ponieważ jest to analiza przeprowadzona na nieporównywalnych korpusach.

toužit		***pragnąć***		***tęsknić***		***marzyć***	
vždycky	751	*rzeczywiście*	13	*strasznie*	2	*dobrze*	3
stále	543	*szczerze*	12	*szczególnie*	2	*niejasno*	2
dlouho	501	*mocno*	8			*naturalnie*	2
vždy	479	*wyraźnie*	8			*nieustannie*	2
také	468	*dużo*	7			*stale*	2

1.6. Krok piąty – gramatyka kognitywna

1.6.1 Słowniki i korpus

Na tym poziomie podjęto próbę odkodowania znaczenia słowa w kontekście konceptualizacji zjawiska nim nazwanego. Analizie poddano czeską jednostkę *mít rád*[21]. W słowniku języka czeskiego (Havránek 1989) *mít rád* jest definiowane jako *pociťovat k někomu náklonnost, lásku, milovat, mít v oblibě* (*czuć do kogoś sympatię miłość, kochać, lubić*). Zgodnie z tymi definicjami słownik czesko-polski (Siatkowski i Basaj 2002) przedstawia polskie ekwiwalenty: *kochać, lubić, przepadać*. Czasowniki te odnoszą się jednak do zupełnie różnych uczuć. Dla osoby polskojęzycznej kombinacja znaczeń „kochać" i „lubić" w ramach jednego wyrażenia jest nieznanym i zaskakującym pojęciem. W korpusie paralelnym InterCorp można znaleźć więcej odpowiedników frazy *mít rád* (*lubić*, *kochać*, *podobać się, uwielbiać, polubić, pokochać, w naszym guście*), jednakże wszystkie należą do dwóch pól semantycznych, odnoszących się do pojęcia „kochać" i „lubić". Niezwykłość tego pojęcia sprawia, iż zarówno rozumienie tej jednostki, jak i jej tłumaczenie na język polski jest bardzo problematyczne. Kwestia ta bywa nierozwiązywalna nawet za pomocą kontekstu, co zostało już wcześniej wspomniane. Na przykład:

(18a) *Mám tě strašně rád, řekl* (Kundera-Valcik_na_rozl).
(18b) *Strasznie cię kocham – rzekł* (Kundera-Valcik_na_rozl).

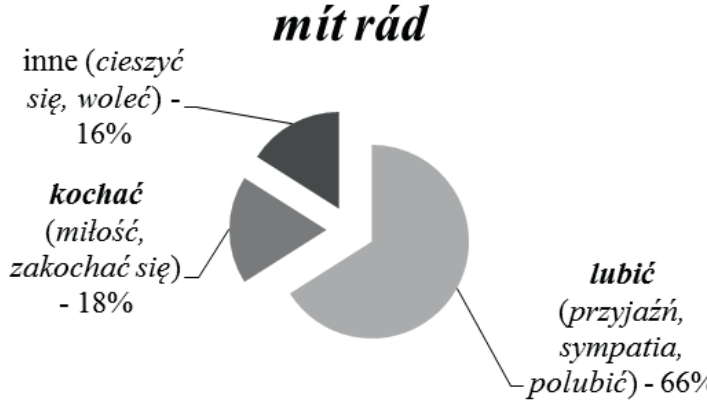
(19a) *Kdybys mě měla ráda, nemohla by ses opičit s tím pitomým jménem* (Grusa-Dotaznik).
(19b) *Gdybyś mnie naprawdę lubiła, nie wygłupiała byś się z tym kretyńskim imieniem* (Grusa-Dotaznik).

[21] Na temat jednostki *mít rád* również w Kaczmarska i Rosen, 2013c.

(20a) *Máš-li mne jen trošku rád, shoď mne z třetího patra, dej mně tu poslední outěchu* (Hasek-OsudyDobrehoVvSV).
(20b) *Jeśli masz dla mnie choć troszkę przyjaźni, zrzuć mnie z trzeciego piętra, udziel mi tej ostatniej pociechy* (Hasek-OsudyDobrehoVvSV).

W kontekstach tych można by użyć wszystkich ekwiwalentów proponowanych przez czesko-polski słownik. Sprawę komplikuje istnienie w języku czeskim jeszcze jednego czasownika tłumaczonego na język polski jako „kochać" – *milovat*. Analiza oparta na materiale z korpusu InterCorp ukazuje, że polski czasownik *kochać* tłumaczony jest na język czeski przeważnie jako *milovat* (ponad 70% poświadczeń). Pojawia się tu jednak dalsza komplikacja dotycząca tłumaczenia jednostki *mít rád*. Rodzimi użytkownicy języka polskiego mają tendencję do odzwierciedlania wzoru z własnego języka w języku obcym (Kaczmarska 2014). W języku polskim czasowniki „lubić" i „kochać" oznaczają odrębne uczucia, a różnica ta polega na jakości, a nie – intensywności. W tej sytuacji osoba polskojęzyczna przypisze znaczenie „kochać" jednostce *milovat*, a znaczenie „lubić" – jednostce *mít rád*. W języku polskim nie istnieje czasownik wyrażający uczucia z pogranicza miłości i sympatii (oraz zawierający jednocześnie oba te dwa uczucia), ponieważ w tym języku nie istnieje takie pojęcie. W rezultacie język polski nie oferuje adekwatnego ekwiwalentu i podczas konfrontacji języka czeskiego i polskiego możemy doświadczyć niezrozumienia (Kaczmarska i Rosen 2014c, 2014a).

W korpusie równoległym InterCorp odnalezionych zostało 2799 poświadczeń jednostki *mít rád*. Z tego 66% przetłumaczone jest na język polski jako jednostka odnosząca się do sympatii (lubić, przyjaźń, sympatia), a 18% – do miłości (kochać, zakochać się)[22].



[22] 16% to inne odpowiedniki.

1.6.2 Ankiety i internet

Zgodnie z dotychczasowymi wynikami badania okazuje się, iż więcej argumentów przemawia za tłumaczeniem czeskiej jednostki *mít rád* jako *lubić* niż jako *kochać*. Ponieważ czasownik ten jest bardzo częsty i bardzo problematyczny, została przeprowadzona ankieta wśród rodzimych użytkowników języka czeskiego[23]. Celem badania było zdefiniowanie znaczenia jednostki *mít rád* na podstawie opozycji z czasownikiem *milovat*. Respondenci byli zapytani również o ewentualne różnice pomiędzy tymi dwoma czeskimi jednostkami, a także poproszeni o podanie obiektów łączących się z tymi czasownikami.

100% respondentów odpowiedziało, iż pomiędzy analizowanymi dwoma jednostkami są różnice semantyczne. Mieli jednak problem z dywersyfikacją obiektów łączących się z tymi jednostkami. Wyniki ankiety zaprezentowano w poniższej tabeli 9.

Tab. 9. Podsumowanie wyników ankiety w Libercu (2013)

	mít rád	***milovat***
definicja	lubić kogoś lub coś, czuć pozytywne emocje	najwyższy stopień miłości, „mít rád" intensywnie, być w głębokiej relacji, czuć miłość, coś więcej niż „mít rád", silne pozytywne uczucie
obiekty	osoba, jedzenie, picie, muzyka, aktywności, piwo, przyroda, rodzice, dziewczyna, życie, zwierzątka domowe…	osoba, aktywność, jedzenie, zwierzątka domowe…

Rezultaty nie były wiążące dla prowadzonej analizy. W tej sytuacji można stwierdzić, że wybór ekwiwalentu zależy od szerszego kontekstu. Do tej pory pracowaliśmy z jednostką *mít rád* w przekonaniu, iż dla osób czeskojęzycznych jest ona jednoznaczna, a dla polskojęzycznych – wieloznaczna, a nawet niejednoznaczna. Jednakże na forach internetowych można znaleźć dużą liczbę wpisów rozważających znaczenie tej jednostki. Rodzimi użytkownicy języka czeskiego dyskutują, co *mít rád* znaczy jako deklaracja[24]. W rezultacie nie możemy stwierdzić, czy problem z tłumaczeniem *mít rád* może być rozwiązany przez szerszy kontekst. Jeżeli dane

[23] Ankieta została przeprowadzona w grudniu 2013 roku w Libercu pośród trzydziestu osób – rodzimych użytkowników języka czeskiego, studentów Technicznego Uniwersytetu w Libercu (Wydział Przyrodniczo-Humanistyczny i Pedagogiczny).

[24] Uczestnicy for internetowych nie mają wątpliwości co do znaczenia czasownika *milovat* w tej samej pozycji: (http://diskuse.doktorka.cz/mit-rad-zamilovat-se-milovat/, http://www.poradte.cz/spolecnost/21684-milovat-nebo-mit-rad.html, http://janajerabkova.blog.idnes.cz/c/194377/Milovat-nebo-mit-rad.html).

pojęcie nie istnieje w języku docelowym, próba odtworzenia go w tym języku może być związana ze stratą części znaczenia lub jego znaczną modyfikacją.

2. Wnioski i perspektywy

Wykorzystanie korpusu w procesie ustalania ekwiwalentów wydaje się oczywiste i konieczne. Korpus równoległy umożliwia ustalenie klasterów ekwiwalentów, co ma istotne znaczenie dla kolejnych etapów analizy. Chociaż korpus paralelny może wspomóc rozwój analiz komparatywnych, badaczom często przychodzi się zmierzyć z niekompatybilnymi narzędziami[25].

Ponieważ uznano, że Word Sketches to program obiecujący dla prowadzonych badań, przygotowano go dla polskiej części korpusu InterCorp, a Word Skeches dla czeskiej części InterCorpu jest w fazie przygotowań[26]. Jest nadzieja, że uruchomienie tego programu zarówno dla polskiej, jak i czeskiej części InterCorpu będzie punktem zwrotnym nie tylko przy budowie tego algorytmu, ale w badaniach kontrastywnych czesko-polskich w ogóle.

Poziom analizy związany z gramatyką przypadków głębokich nie przyniósł oczekiwanych rezultatów. Metoda wykorzystująca tę teorię lingwistyczną musi być przetestowana na większej liczbie czasowników. Oczywistym też jest, iż opracowanie to wymaga głębszej analizy kognitywnej większości problematycznych jednostek, choć z drugiej strony metody kognitywne bardzo trudno implementować do algorytmu, ponieważ przeprowadzane są w większości manualnie[27].

Przeprowadzone badania ukazują też jednoznacznie problem nieobecności jakiegoś pojęcia w jednym z badanych języków. Tłumaczenie słowa (czy frazy) owego pojęcia nazywającego prowadzi często do arbitralnej decyzji tłumacza.

Miejmy nadzieję, że w przyszłości algorytm ten będzie współpracował z programami do tłumaczenia maszynowego, dlatego uzupełnieniem analizy manualnej będą również eksperymentalne próby stochastycznego modelowania wyboru ekwiwalentu na podstawie analizy kontekstu, które zostały przedstawione w odrębnym artykule (zob. Kaczmarska, Rosen, Hana i Hladká 2015).

---

[25] Napotykano na ten sam problem przy korzystaniu z korpusu jednojęzycznego. Narzędzie Word Sketches dostępne jest dla SYN (Czeski Korpus Narodowy). Dla języka polskiego korpus porównywalny stanowi NKJP (Narodowy Korpus Języka Polskiego), ale dla niego nie można zastosować Word Sketches. Poza tym korpusy narodowe czeski i polski mają inne funkcje statystyczne, co jeszcze bardziej komplikuje analizy komparatywne.

[26] Niestety, narzędzie to nie jest dostępne dla zewnętrznych użytkowników.

[27] W wypadkach problemowych można odnieść się także do eksplikacji i naturalnego metajęzyka semantycznego (Wierzbicka 1980, 2001) lub skonstruować skalę intensywności właściwości wyrażanej przez dany czasownik (Mikołajczuk 1997, 1999; Bratman 1987).

Bibliografia

Bratman, M.E. (1987). *Intentions, Plans, and Practical Reason*. Massachusetts: Harvard University Press.

Čermák, F., Rosen A. (2012). The case of InterCorp, a multilingual parallel corpus, *International Journal of Corpus Linguistics* 13(3), 411–427.

Čermáková, A. (2009). *Valence českých substantive*. Praha: Nakladatelství Lidové noviny.

Daneš, F., Hlavsa. Z. (1987). *Větné vzorce v češtině*. Praha: Academia.

Ebeling, J., Ebeling. S.O. (2013). *Patterns in contrast*. John Benjamins Publishing Company.

Fillmore, Ch.J. (1968). The Case for Case. W: E. Bach i R.T. Harms (red.). *Universals in Linguistic Theory* (1–88). New York: Holt, Rinehart, and Winston.

Fisher, H. (2005). *Anatomia miłości*. Tł. Joteł (Jerzy Łoziński).Warszawa: Rebis.

Greń, Z., Rytel-Kuc, D. (1991). *Wykorzystanie przekładów literackich w pracy nad dwujęzycznym słownikiem walencyjnym*. W: H. Běličová, G. Nieszczimienko, Z. Rudnik-Karwatowa (red.). *Problemy teoretyczno-metodologiczne badań konfrontatywnych języków słowiańskich* (69–78). Warszawa: Instytut Słowianoznawstwa Polskiej Akademii Nauk.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London: Arnold.

Han, A.L., Lu, Y., Wong, D.F., Chao, L.S., He, L., Junwen, X. (2013). Quality Estimation for Machine Translation Using the Joint Method of Evaluation Criteria and Statistical Modeling. W: *Proceedings of the Eighth Workshop on Statistical Machine Translation*, 365–372. Association for Computational Linguistics.

Hatfield, E., Rapson, R.L. (1993). *Love and attachment processes*. W: M. Lewis, J.M. Haviland (red.) *Handbook of emotions* (595–604). New York: Guilford Press, 1993. Pozyskano 15.05.2014 z: http://www.elainehatfield.com/uploads/3/2/2/5/3225640/48._hatfield__rapson_1993.pdf.

Havránek, B. (red.). (1989). *Slovník spisovného jazyka českého*. Praha: Academia.

Hunston, S., Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins Publishing Company.

Jirásek, K. (2011). Využití paralelního korpusu InterCorp k získávání ekvivalentů pro chorvatsko-český slovník. W: F. Čermák (red.). *Korpusová lingvistika Praha 2011: 1 – InterCorp* (45–55). Praha: Nakladatelství Lidové noviny, 2011.

Kaczmarska, E. (2001). Badanie struktury walencyjnej czeskich i polskich predykatów posiadających pozycję Experiencera. *Studia z Filologii Polskiej i Słowiańskiej* 37, 177–187.

Kaczmarska, E. (2010). *Analiza zdolności konotacyjnych polskich i czeskich predykatów odnoszących się do strachu, złości i wstydu*. W: J. Goszczyńska, Z. Greń (red.). *Res slavisticae* (135–153). Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.

Kaczmarska E. (2014). *Czeskie czasowniki oznaczające stany psychiczne – sposoby ustalania polskich ekwiwalentów na podstawie korpusu równoległego InterCorp*.

W: A. Stolarczyk-Gembiak, M. Woźnicka (red.). *Zbliżenia: językoznawstwo – literaturoznawstwo – translatologia* (45–55). Konin: Państwowa Wyższa Szkoła Zawodowa w Koninie.

Kaczmarska, E., Rosen, A. (2014c). *Czego nie można wyrazić w języku polskim, czyli o leksykalnych w nim brakach*. Polonica 334, 53–66.

Kaczmarska, E., Rosen, A. (2013b). Między znaczeniem leksykalnym a walencją – próba opracowania metody ekstrakcji ekwiwalentów na podstawie korpusu równoległego. *Studia z Filologii Polskiej i Słowiańskiej* 48, 103–121.

Kaczmarska, E., Rosen, A. (2014a). *Niedosłowności w dialogu czesko-polskim* (w druku). Referat wygłoszony na konferencji *Język trzeciego tysiąclecia: (Nie)dosłowność*, Kraków 13–14.03.2014.

Kaczmarska, E., Rosen, A. (2014b). *Praktyczny przewodnik po korpusie równoległym InterCorp*. W: M. Hebal-Jezierska (red.). *Praktyczny przewodnik po korpusach języków słowiańskich*, (207–231). Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.

Kaczmarska, E., Rosen, A., Hana, J., Hladká, B. (2015). *Syntactico-semantic analysis of arguments as a method for establishing equivalents of Czech and Polish verbs expressing mental states*. Prace Filologiczne LXVII, 151–174.

Kędzia, P., Piasecki, M., Kocoń, J., Indyka-Piasecka, A. (2014). *Distributionally Extended Network-Based Word Sense Disambiguation in Semantic Clustering of Polish Texts*. W: *IERI Procedia (International Conference on Future Information Engineering)* 10, 38–44. DOI: 10.1016/j.jeri.2014.09.073.

Korytkowska, M. (1984). *Kategoria przypadka semantycznego (na materiale języka polskiego, bułgarskiego i serbsko-chorwackiego)*. W: V. Koseska-Toszewa, M. Korytkowska (red.). *Studia konfrontatywne polsko-południowosłowiańskie* (11–38). Wrocław: Zakład Narodowy im. Ossolińskich.

Korytkowska, M. (1993). *O konfrontatywnym opisie predykatorów bułgarskich i polskich (na przykładzie jednostek otwierających miejsce dla argumentu o wartości Experiencer)*. W: V. Koseska-Toszewa, M. Korytkowska (red.). *Studia gramatyczne bułgarsko-polskie* V–VI (121–150). Warszawa: Slawistyczny Ośrodek Wydawniczy.

Korytkowska, M. (1992). *Typy pozycji predykatowo–argumentowych. Gramatyka konfrontatywna bułgarsko-polska*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Langacker, R. (1987). *Foundations of Cognitive Grammar, t. 1, Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, R. (1991). *Foundations of Cognitive Grammar, t. 2. Descriptive Application*. Stanford: Stanford University Press.

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.

Lewandowska-Tomaszczyk, B. (1984). *Conceptual Analysis, Linguistic Meaning, and Verbal Interaction*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Lewandowska-Tomaszczyk, B. (2013). *Komunikacja i konstruowanie znaczeń w przekładzie*. Referat prezentowany na konferencji *Zbliżenia: językoznawstwo – translatoryka – literaturoznawstwo*, Konin 13–14.11.2013.

Linde-Usiekniewicz, J., Derwojedowa, M., Zawisławska, M., (2008). *Verbal Aspect and the Frame Elements in the FrameNet for Polish*. Pozyskano z: http://www2.polon.uw.edu.pl/derwojed/wp-content/uploads/2012/07/euralex_linde_derwojedowa_zawislawska.pdf.

Mikołajczuk, A. (1999). *Gniew we współczesnym języku polskim. Analiza semantyczna*. Warszawa: Wydawnictwo Energeia.

Mikołajczuk, A. (1997). *Pole semantyczne 'gniewu' w polszczyźnie (Analiza leksemów: gniew, oburzenie, złość, irytacja)*. W: R. Grzegorczykowa Z. Zaron (red.). *Semantyczna struktura słownictwa i wypowiedzi* (149–171). Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

Młodzki, R., Kopeć, M. Przepiórkowski, A. (2012). Word Sense Disambiguation in the National Corpus Of Polish. *Philological Studies* (*Prace Filologiczne*) LXIII: 155–166.

Och, F.J., Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29 / 1, 19–51.

Pajdzińska, A. (1999). *Jak mówimy o uczuciach? Poprzez analizę frazeologizmów do językowego obrazu świata*. W: J. Bartmiński (red.). *Językowy obraz świata* (87–107). Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Pęzik, P. (2012). *Wyszukiwarka PELCRA dla danych NKJP*. W: A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (reds.). *Narodowy Korpus Języka Polskiego* (253–279). Wydawnictwo PWN .

Rytel, D. (1989). Wybrane problemy opisu walencyjnego języka. *Studia z Filologii Polskiej i Słowiańskiej* 26, 237–247.

Rytel-Kuc, D. (red.). (1991). *Walencja czasownika a problemy leksykografii dwujęzycznej*. Wrocław: Zakład Narodowy im. Ossolińskich.

Siatkowski, J., Basaj, M. (2002). *Słownik czesko-polski*. Warszawa: Wiedza Powszechna.

Skoumalová, H. (2008). *Extracting dictionaries from parallel corpora*. W: *Proceedings of The Third Baltic Conference on Human Language Technologies* (297–301). Kaunas: Vytautas Magnus University.

Tian, L., Wong, D.F., Chao, L.S., Oliveira, F. (2014). A Relationship: Word Alignment, Phrase Table, and Translation Quality. *The Scientific World Journal*. Hindawi Publishing Corporation. Pozyskano z: http://dx.doi.org/10.1155/2014/438106.

Tian, L., Wong, D.F., Chao, L.S. (2010). An Improvement of Translation Quality with Adding Key-Words in Parallel Corpus. W: *Machine Learning and Cybernetics* (ICMLC) t. 3, 1273 – 1278. DOI: 10.1109/ICMLC.2010.5580888.

Urbańczyk-Adach, N. (2011). *Wariantywność walencji czeskiego czasownika*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Zawisławska M. (1998). The meaning of English verb see and its Polish equivalent *widzieć* in a perspective of frame semantics. W: W. Teubert, E. Tognini Bonelli, N. Volz (red.) *TERLI. Trans-European Language Resources Infrastructure*. Proceedings of the Third European Seminar „Translation Equivalence", Montecatini Terme, Italy October 16–18, 1997. Mannheim: TELRI Association e.V. Institut fur deutsche Sprache, The Tuscan Word Centre.

Korpusy

Czech National Corpus – InterCorp. Institute of the Czech National Corpus. Available online at http://www.korpus.cz.

National Corpus of Polish. Available online at: http://nkjp.pl.

In search of the meaning of verbs expressing mental states. An analysis of Czech verbs and their Polish equivalents – an attempt at implementing different linguistic theories (valency, case grammar, pattern grammar, cognitive linguistics)

s u m m a r y

The analysis is focused on Czech polysemous verbs expressing mental states. We have tried to build an algorithm establishing their Polish equivalents and we tested if different linguistic theories can lead us to the closest equivalents. The research includes automatic excerption of chosen verbs (with aligned segments) from a parallel corpus – InterCorp. The segments were analysed manually. We checked how a given verb was translated and what kinds of collocations and arguments it had. Subsequently, we applied methods that were specific for each linguistics approach. We also used Word Sketches, which have proved to be a fundamental tool. Our analysis shows the advantages of these theories, but also deficiencies in the proposed algorithm that must be either corrected or developed.