

Elżbieta Kaczmarska

Instytut Sławistyki Zachodniej i Południowej

Wydział Polonistyki

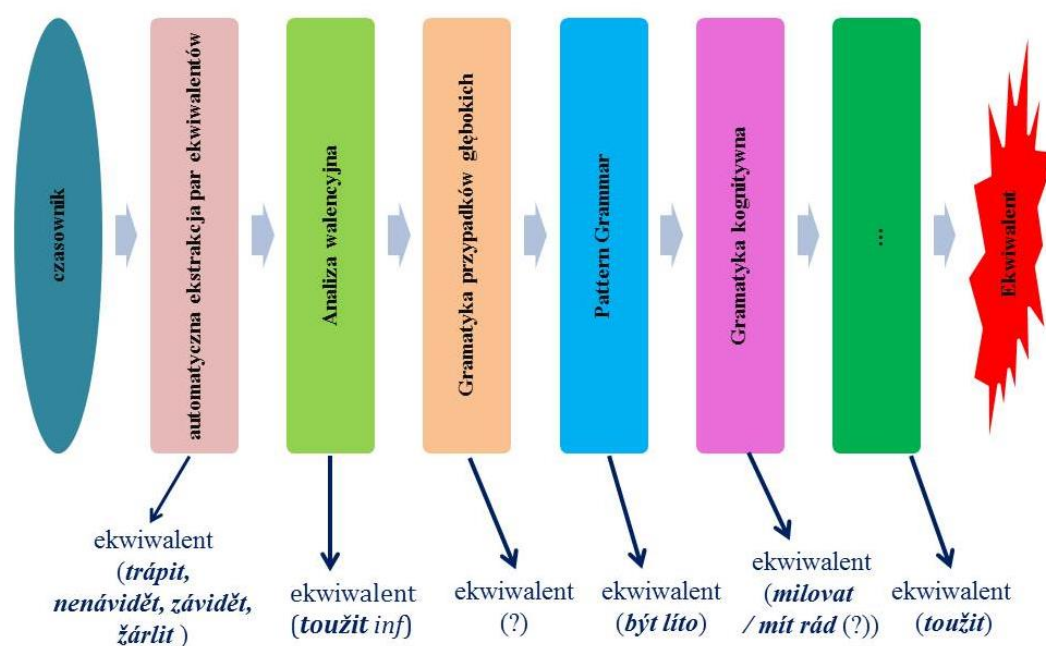
Uniwersytet Warszawski

*Metody ustalania ekwiwalentów czasowników wyrażających stany
emocjonalne w przekładzie czesko-polskim (PROJEKT ALGORYTMU)*

4 PROJEKT ALGORYTMU

Ustalenie ekwiwalentów dla omawianych czeskich czasowników było częścią większego projektu, w ramach którego opracowywany jest algorytm ułatwiający wyszukiwanie odpowiedników czasowników wyrażających różne uczucia i emocje na podstawie danych z korpusu równoległego InterCorp (Kaczmarska 2015A)¹. Implementując metody różnych teorii językoznawczych (zob. Obraz 11), sprawdzamy zależność znaczenia od charakterystyki składniowej (Levin 1993) i tym samym, na podstawie schematów składniowych, ustalamy najbliższy ekwiwalent (odnosząc się również do struktury składniowej potencjalnego odpowiednika).

¹ 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.



Obraz 7: Schemat wypracowywanego algorytmu wspomagającego ustalenie ekwiwalentów.

Jak zostało to przedstawione, cały algorytm składa się z kilku etapów analizy. Czasownik poddawany takiemu badaniu nie musi przechodzić przez wszystkie etapy; optymalny ekwiwalent może być odnaleziony na każdym poziomie analizy.

Pierwszym krokiem jest automatyczna ekstrakcja par ekwiwalentów. Badania pilotażowe ukazały, iż wiele czasowników już na tym etapie odnajduje trafne ekwiwalenty; co ciekawe – są to na ogół czasowniki wyrażające negatywne odczucia (Kaczmarek 2013).

Drugim krokiem jest ustalenie wymagań walencyjnych. Zakładamy, iż zbieżność struktur składniowych i podobieństwo semantyczne obiektów łączących się z danymi czasownikami może być kluczem do wskazania ekwiwalentu na tym etapie. Niektóre z czasowników odnajdą trafne ekwiwalenty tylko dla niektórych swoich znaczeń. Tak się dzieje w przypadku czasownika *toužit*.² Czasowniki, które nie odnalazły swoich ekwiwalentów na tym etapie przechodzą do następnego.

² Na podstawie wyników analizy walencyjnej i łączliwości udało się nam ustalić, iż ekwiwalentem czasownika *toužit* łączącego się z bezokolicznikiem jest polski czasownik *pragnąć* łączący się również z bezokolicznikiem (Kaczmarek i Rosen 2013).

Trzecim krokiem, nieprzynoszącym jak do tej pory oczekiwanych rezultatów, jest identyfikowanie przypadków głębokich – roli, jakie pełnią elementy wiążące się z danym czasownikiem.³

Czwarty krok to wykorzystanie metod teorii Pattern Grammar (Ebeling i Ebeling 2013) (Hunston i Francis 2000). Ten etap ma w założeniu 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. Sprawdziło się to m.in. w przypadku frazy *být lito*. Jeżeli jednostka *být lito* łączy się z dwoma frazami nominalnymi (w celowniku i dopełniaczu), wówczas taki jej wzorzec odpowiada polskiemu ekwiwalentowi *zał*. Jeżeli natomiast *být lito* łączy się tylko z celownikową frazą nominalną (oraz ewentualnie z elementem *to*), wzorzec ten odpowiada znaczeniu polskiej jednostki (*być*) *przykro* (Kaczmarska 2015A).

Kolejnym krokiem jest etap wykorzystania założeń gramatyki kognitywnej. Na tym poziomie próbujemy odkodować znaczenie słowa w kontekście konceptualizacji zjawiska nim nazwanego. W badaniach pilotażowych analizowaliśmy jednostki *mít rád* i *milovat*. O ile poradziliśmy sobie z czasownikiem *milovat*, o tyle identyfikacja znaczenia jednostki *mít rád* i jej przekład na polski wciąż pozostaje problematyczne. Metody kognitywne jednak bardzo trudno implementować do algorytmu, ponieważ przeprowadzane są w większości manualnie.⁴

Ostatnie okienko to miejsce na rozwiązania przyszłościowe takie, jak np. wykorzystanie narzędzia WordSketch dla obu języków na tym samym korpusie, czy zaawansowane badania stochastyczne. Wierzymy, że dzięki nim, na podstawie badania kontekstu i rozróżnienia łączliwości z różnymi obiektami, uda nam się dopasować ekwiwalenty do wszystkich znaczeń czasownika *toužit* (Kaczmarska, Rosen i Hana i inni 2015), a także wszystkich innych, które nie znajdują swoich ekwiwalentów na wcześniejszych poziomach algorytmu.

³ Etap ten nie zostanie jednak usunięty z finalnej wersji algorytmu. Badane przez nas jednostki (wyrażające różne emocje i uczucia) są jednolite pod względem łączliwości z argumentami o pewnych wartościach – zawsze znajdziemy tu przeżywającego, czy źródło i na ten podstawie nie dokonamy różnicowania znaczenia. Algorytm może być jednak wykorzystany w badaniu innych grup czasowników, gdzie rola semantyczna oryginału i ekwiwalentu może być istotna przy różnicowaniu znaczenia.

⁴ W przypadkach problemowych możemy 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).

4.1. Walencja

Celem analizy na tym poziomie badania jest odpowiedź, czy wymogi walencyjne⁵ mogą ułatwić identyfikację polskiego ekwiwalentu danego czeskiego czasownika. W tym celu przeprowadziliśmy badanie pilotażowe czasownika *toužit* (Kaczmarska i Rosen 2013). Założyliśmy, iż w niektórych przypadkach ekwiwalent może być ustalony na podstawie zbieżności wymagań walencyjnych (Levin 1993). Kolejno przeprowadziliśmy analizę manualną wybranych czasowników (wraz z wiązanymi segmentami). W każdym segmencie określiliśmy, jak dany czeski czasownik został przetłumaczony na język polski, ile argumentów wiąże i jakiego rodzaju argumentu to są⁶ oraz w jaki sposób są z danym czasownikiem połączone.⁷ Skontrolowaliśmy także, jakie argumenty łączą się z polskimi ekwiwalentami oraz przeprowadziliśmy syntaktyczną i semantyczną analizę argumentów wiążących się z czeskimi czasownikami i ich polskimi ekwiwalentami. Oczekiwaliśmy, 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. Badanie to ukazuje, iż walencja może wpłynąć na wybór polskiego ekwiwalentu. W przypadku analizowanego czasownika optymalny ekwiwalent identyfikowany jest jednak tylko dla struktury *toužit* + bezokolicznik:

toužit + *inf* → *pragnąć*⁸ + *inf*.

⁵ 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)).

⁶ Np. fraza nominalna – jeśli tak, to jakiego typu: obiekt realny, abstrakcyjny, nazywający istotę ludzką.

⁷ 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 zaobserwowaliśmy podczas syntaktycznej analizy manualnej. Tylko pozornie jest to pomieszczenie dwóch płaszczyzn analizy. Praktycznie polegało to na dokładniejszym „tagowaniu“ obiektu łączącego się z przyimkiem „po“.

⁸ Czasownik *chciē* jest klasyfikowany jako synonim *pragnāc*. Uznajemy, iż różnica pomiędzy nimi polega na intensywności uczucia. Podobnie pozostałe jednostki, podkreślone w tabeli 3.

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

Tabela 3. Polskie ekwiwalenty *toužit* + bezokolicznik.

W pozostałych grupach uzyskane rezultaty nie są rozstrzygające⁹. Szczególnie struktury z obiektem abstrakcyjnym okazały się niezwykle problematyczne i wskazały na konieczność głębszej analizy semantycznej obiektów.¹⁰

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.

4.2. Gramatyka przypadków głębokich

⁹ Szczegółowe wyniki opublikowane zostały w (Kaczmarska i Rosen 2013).

¹⁰ Przeprowadziliśmy test z dwoma obiektami abstrakcyjnymi – *skutečné přátelství / prawdziwa przyjaźń* i *nezapomenuté dobrodružství / niezapomniana przygoda*. Zaobserwowaliśmy, iż oba obiekty łączą się z analizowanym czeskim czasownikiem *toužit po skutečném přátelství / po nezapomenutelném dobrodružství*. Spróbowaliśmy więc oba połączyć 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 2013) (Kaczmarska, Rosen i Hana i inni 2015)

Na tym poziomie ustalamy, czy jednostki kandydujące na ekwiwalenty łączą się z obiektami reprezentującymi te same kategorie ról semantycznych¹¹. 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 Experiencer¹², np.

(5a) *Lucie mne miluje a její zdraví závisí na mé lásce!* (kundera-zert)¹³

(5b) *Lucja 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 opovrživým 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. Ponownie analizie poddamy czeski czasownik *toužit*. W przykładach (8a) – (10b) podkreślone zostały pozostałe argumenty łączące się z badanym czasownikiem, ale nie będące obiektem o wartości Experiencer.

¹¹ Krok ten dotyczy sprawdzania zgodności przypadków semantycznych we frazach z czasownikiem w języku wyjściowym i ekwiwalentem w języku docelowym (Fillmore 1968) (Halliday 1985) (Korytkowska 1984) (Korytkowska 1992) (Korytkowska 1993) (Kaczmarska 2001) (Kaczmarska 2010).

¹² Wyodrębnienie argumentu o wartości Experiencer wymagało ustalenia, jakie predykaty otwierają miejsce dla tego argumentu. Klasa predykatów otwierających miejsce dla argumentu o wartości Exp 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, które zachodzą 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” nie działający świadomie i celowo. Por. Experiencer (Kaczmarska 2001) (Kaczmarska 2010) (Korytkowska 1992) (Korytkowska 1993) Proživatel (Karlík, Nekula, Rusínová 2002), Senser (Halliday 1985).

¹³ Przykłady pochodzą z korpusu InterCorp; informacja bibliograficzna umieszczona jest na końcu każdej frazy w nawiasie.

(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 zostało już wcześniej wspomniane, 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) *Podarilo 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.

4.3. Pattern Grammar

Ten poziom analizy w założeniu ma pozwolić sprawdzić, czy istnieje zależność pomiędzy znaczeniem wyrazu a otoczeniem, w jakim występuje¹⁴. Hunston i Francis twierdzą, że jeżeli jakieś słowo jest wieloznaczne, a jednocześnie pojawia się w kilku wzorcach, to każdy wzorec pojawi się z jednym z jego znaczeń częściej niż z pozostałymi, czyli dany wzorec 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: 20)

Czasowniki wyrażające stany psychiczne są w większości polisemiczne. Poprzez śledzenie i identyfikację ich wzorców, mamy więc nadzieję połączyć konkretne znaczenie z typem wzorca (rozumianym jako powtarzalna kombinacja słów).¹⁵

Podjęliśmy próbę ustalenia, czy w rzeczywistości istnieje taka powtarzalność w poświadczeniach korpusowych (por. Ebeling and Ebeling 2013). Ręczna analiza oparta na materiale z korpusu InterCorp ukazała między innymi dwa wzorce czeskiej jednostki *být líto* (być przykro, żalować) związanych z dwoma znaczeniami. Jeżeli jednostka *být líto* łączy się z dwoma frazami nominalnymi (w celowniku i dopełniaczu), wówczas taki jej wzorec odpowiada polskiemu ekwiwalentowi *žal*. Jeżeli natomiast *být líto* łączy się tylko z celownikową frazą nominalną (oraz ewentualnie z elementem *to*), wzorec ten odpowiada znaczeniu polskiej jednostki (*być*) *przykro*. Ilustruje to poniższa tabelka (4).

¹⁴ Dokonana zostanie analiza wzorców, w jakich pojawiają się badane jednostki – Pattern Grammar (Francis i in. 1996; Hunston i Francis 2000; Ebeling i Ebeling 2013).

¹⁵ “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 and Francis 2000: 37)

<i>žal</i>	<i>(być) przykro</i>
(12a) <i>Jak mi ho bylo líto!</i>	(15a) <i>Pak mi je líto.</i>
(12b) <i>Jakže mi go bylo žal!</i>	(15b) <i>Wobec tego, przykro mi!</i>
(13a) <i>Je mi ho samozřejmě líto.</i>	(16a) <i>Potom nám to bylo oběma líto.</i>
(13b) <i>Jest mi go oczywiście žal...</i>	(16b) <i>Potem nam obu bylo przykro.</i>
(14a) <i>Přišlo mi jí prostě líto.</i>	(17a) <i>...nabídne mi sisinku a já si vezmu, protože by mu bylo líto, kdybych si nevzala...</i>
(14b) <i>Po prostu zrobiło mi się jej žal.</i>	(17b) <i>...zaprasza mnie na cuksa i ja biorę, bo byłoby mu przykro, gdybym nie wzięła...</i>
<i>být líto + NP_{DAT} + NP_{GEN} = žal</i>	<i>být líto + NP_{DAT} + to / Ø = (być) przykro</i>

Tabela 4. Wzorce czeskiej jednostki *být líto* (*žal*, *być przykro*).

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

4.4. Word Sketches

Word Sketches to nie jest teorią językoznawczą. To program umożliwiający przedstawienie obrazu kolokacyjnego danego słowa również z odniesieniem do gramatyki.¹⁶ 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, dokonujemy próby ponownej analizy czeskiego czasownika *toužit*.

Badanie zostało przeprowadzone z oparciem o czesko-polską część korpusu równoległego InterCorp. Na początku staraliśmy się potwierdzić tezę o ekwiwalencji *toužit* + bezokolicznik i *pragnąć* + bezokolicznik. Sprawdziliśmy, który z ekwiwalentów łączy się z

¹⁶ “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>

bezokolicznikiem. Materiał korpusowy ukazał, iż polskie czasowniki *marzyć* i *tęsknić* nie łączą się z bezokolicznikiem. Rezultaty prezentowane są w tabelce (5); zawiera ona jedynie 10 najczęstszych obiektów.

<i>toužit inf</i>		<i>pragnąc inf</i>		<i>* marzyć inf</i>	<i>* tęsknić inf</i>
<i>post_inf</i>	17405	<i>post_inf</i>	6800		
<i>mít</i>	926	<i>podziękować</i>	805		
<i>stát</i>	864	<i>podkreślić</i>	598		
<i>poznat</i>	382	<i>pogratulować</i>	379		
<i>vidět</i>	346	<i>wyrazić</i>	391		
<i>vrátit</i>	333	<i>zwrócić</i>	319		
<i>hrát</i>	333	<i>przypomnieć</i>	165		
<i>získat</i>	332	<i>powiedzieć</i>	386		
<i>dostat</i>	311	<i>zauważyć</i>	97		
<i>vyhrát</i>	285	<i>rozpocząć</i>	73		
<i>žít</i>	177	<i>poruszyć</i>	58		

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

W tym momencie napotykamy 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. wyniki w tabeli (6) – poniżej]. Obecnie wykorzystanie narzędzi Word Sketches dla języka czeskiego jest niemożliwe z użyciem InterCorpu. Nie jest możliwe również wykorzystanie Narodowego Korpusu Języka Polskiego (dalej – NKJP); 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.

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

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

Wyniki w tabeli (6) są niejednoznaczne, postanowiliśmy więc przeprowadzić eksperymentalne badanie, wykorzystując Word Sketches dla języka polskiego (InterCorp), kolokator – również dla języka polskiego (NKJP)¹⁷ oraz Word Sketches dla języka czeskiego (CNK). Wyniki tej analizy przedstawia tabela (7).

¹⁷ Kolokator dostępny na stronie: <http://www.nkjp.uni.lodz.pl/collocations.jsp>

Word Sketches (InterCorp)	kolokator ¹⁸ (NKJP)	Word Sketches (CNK)
<i>pragnąć</i>	809 <i>pragnąć + Gen</i>	<i>toužit po</i>
<i>has_gen_obj</i>		<i>post_po</i>
<i>co</i>	76 <i>on</i>	1460 <i>dítě</i>
<i>to</i>	52 <i>człowiek</i>	163 <i>láska</i>
<i>Europa</i>	34 <i>ty</i>	143 <i>návrat</i>
<i>zachęcić</i>	26 <i>życie</i>	110 <i>úspěch</i>
<i>strona</i>	16 <i>coś</i>	107 <i>vítězství</i>
<i>coś</i>	15 <i>bóg</i>	63 <i>změna</i>
<i>powód</i>	14 <i>kobieta</i>	60 <i>život</i>
<i>region</i>	14 <i>dziecko</i>	57 <i>medaile</i>
<i>śmierć</i>	13 <i>świat</i>	50 <i>pomsta</i>
<i>demokracja</i>	12 <i>nic</i>	47 <i>klid</i>
		23752

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

Rezultaty tego badania mają dla nas znaczenie wyłącznie poznawcze, a samego testu nie możemy implementować do algorytmu wyszukiwania ekwiwalentów. Program Word Sketches wydaje się obiecującym narzędziem dla naszej analizy. Optymalną sytuacją byłoby, gdyby analiza była przeprowadzona na podstawie jednego korpusu i jednobrzmiących tekstów. Dlatego narzędzie, nad którym aktualnie pracujemy, współpracować będzie zarówno z czeską, jak i polską częścią InterCorpu. Oczekujemy, iż będziemy w stanie analizować nie tylko obiekty wyrażone frazami nominalnymi, ale także przysłowki łączące się z danym czasownikiem. Poniższa tabelka (8) przedstawia rozkład przysłówków w odniesieniu do interesujących nas jednostek.¹⁹

¹⁸ W przypadku kolokatora wyniki zostały zliczone ręcznie, a takiej analizie został poddany tylko jeden polski ekwiwalent – *pragnąć*.

¹⁹ Ponownie jest to jednak badanie testowe; wyników nie możemy brać pod uwagę we wnioskach czy podsumowaniu, ponieważ jest to analiza dokonana na nieporównywalnych korpusach.

<i>toužit</i>		<i>pragnąć</i>		<i>tęsknić</i>		<i>marzyć</i>	
<i>hodně</i>	1099	<i>bardzo</i>	136	<i>bardzo</i>	34	<i>jedynie</i>	6
<i>moc</i>	1085	<i>gorąco</i>	40	<i>ogromnie</i>	3	<i>często</i>	5
<i>tak</i>	783	<i>jedynie</i>	19	<i>niesamowicie</i>	2	<i>bardzo</i>	5
<i>už</i>	778	<i>jednocześnie</i>	16	<i>okropnie</i>	2	<i>próżno</i>	4
<i>tolik</i>	773	<i>rozpaczliwie</i>	15	<i>straszliwie</i>	2	<i>długo</i>	4
<i>vždycky</i>	751	<i>rzeczywiście</i>	13	<i>strasznie</i>	2	<i>dobrze</i>	3
<i>stále</i>	543	<i>szczerze</i>	12	<i>szczególnie</i>	2	<i>niejasno</i>	2
<i>dlouho</i>	501	<i>mocno</i>	8			<i>naturalnie</i>	2
<i>vždy</i>	479	<i>wyraźnie</i>	8			<i>nieustannie</i>	2
<i>také</i>	468	<i>dużo</i>	7			<i>stale</i>	2

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

4.5. Cognitive Grammar

Na tym poziomie spróbujemy odkodować znaczenie słowa w kontekście konceptualizacji zjawiska nim nazwanego²⁰. Poddamy analizie czeską jednostkę *mít rád*.²¹ W słowniku języka czeskiego (Havránek 1989) *mít rád* jest definiowane jako *pociťovat k někomu náklonnost, lásku, milovat, mít v oblíbě* (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 to zupełnie różnych uczuć. Dla osoby polskojęzycznej kombinacja znaczeń ‘kochać’ i ‘lubić’ w ramach jednego wyrażenia jest nieznanym i zaskakującym pojęciem. W korpusie paralelnym InterCorp możemy 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

²⁰ Wykorzystanie elementów gramatyki kognitywnej za: Langacker 1987, 1991, 2008; Taylor 2002; Mikołajczuk 1997, 1999.

²¹ Na temat jednostki *mít rád* również w: Kaczmarska i Rosen, 2013a.

sprawia, iż zarówno rozumienie tej jednostki, jak i jej tłumaczenie na język polski jest bardzo problematyczne; problem ten bywa nierozwiązywalny nawet przez kontekst, co zostało już wcześniej wspomniane. Np.:

(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)

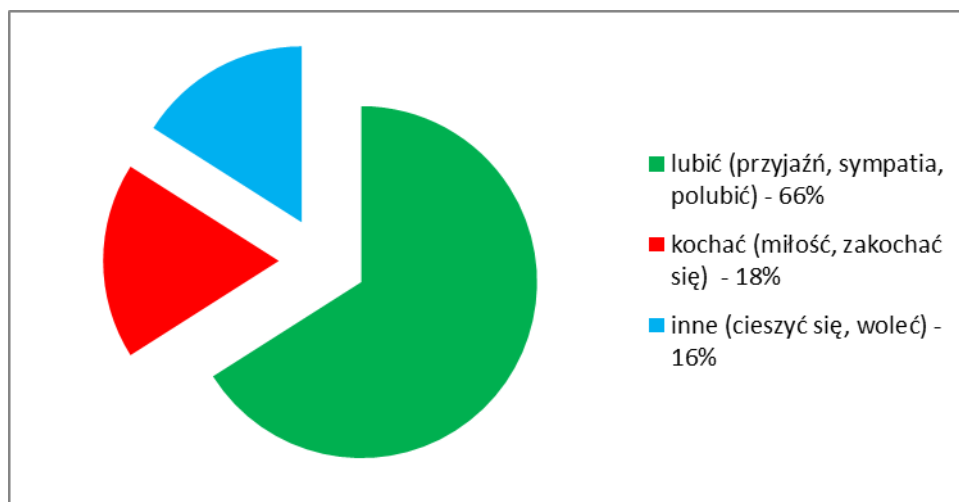
(19b) *Gdybyś mnie naprawdę lubiła, nie wygłupiała byś się z tym kretyńskim imieniem.*
(Grusa-Dotaznik)

(20a) *Máš-li mne jen trošku rád, shod' 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 moglibyśmy 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 (Kaczmarek 2013). W języku polskim ‘lubić’ i ‘kochać’ oznaczają różne uczucia, a różnica ta polega na jakości, a nie na 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 nie istnieje takie pojęcie w tym języku. W rezultacie język polski nie oferuje adekwatnego ekwiwalentu, a podczas konfrontacji języka czeskiego i polskiego możemy doświadczyć niezrozumienia (Kaczmarek i Rosen 2013a, 2014a).

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



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 prezentuje poniższa tabela (9).

²² 16% to inne odpowiedniki.

²³ Ankieta została przeprowadzona w Libercu (Republika Czeska), w grudniu 2013 (30 respondentów w wieku od 19 do 57 lat).

	<i>mít rád</i>	<i>milovat</i>
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...

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

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

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>)